DOCUMENT RESUME

ED 192 739 IR 008 804

AUTHOR Siegel, Arthur I.: And Others
TITLE Application of Computer Simulation Techniques in Military Exercise Control System Development: 1. NETMAN Model Sensitivity Test and Validation. Technical Report 407.
INSTITUTION Applied Psychological Services, Inc., Wayne, Pa.
SPONS AGENCY Army Research Inst. for the Behavioral and Social Sciences, Alexandria, Va.
PUB DATE Sep 79
CONTRACT MDA903-78-C-2013
NOTE 115p.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Computers; Management Systems; *Mathematical Models; *Military Training; *Simulation; Systems Analysis

ABSTRACT

The methods, procedures, and results of a sensitivity test on NETMAN, a stochastic digital computer model designed to simulate the information collection and scoring control systems used in tactical warfare training exercises, are described. NETMAN enables its users to obtain information such as personnel requirements, training requirements, and workload alternatives in such systems as the Marine Corps' Tactical Warfare Simulation, Evaluation, and Analysis System (TWSEAS). The model had received preliminary sensitivity testing, and a more thorough program was required. Consequently the Army Research Institute instituted a program to complete a formal model sensitivity test and validation, and define the role of such exercise control models in system design and redesign. This report summarizes the first year's effort and provides the model validation results, including selection and collection of criterion data from actual military exercises and the comparison of model results against these measures. Conclusions indicate that the model can be used with increased confidence to predict the performance of training exercise control systems up to battalion level. (Author/LLS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED192739

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Technical Report 407

APPLICATION OF COMPUTER SIMULATION TECHNIQUES IN MILITARY EXERCISE CONTROL SYSTEM DEVELOPMENT: I. NETMAN Model Sensitivity Test and Validation

Arthur I. Siegel, W. Rick Leahy
and J. Jay Wolf
Applied Psychological Services, Inc.

and

Thomas G. Ryan
U.S. Army Research Institute for the
Behavioral and Social Sciences

Robert T. Root, Team Leader

Submitted by:
Frank J. Harris, Chief
ENGAGEMENT SIMULATION TECHNICAL AREA

Approved by:

Milton S. Katz, Ph.D., Acting Director
ORGANIZATIONS AND SYSTEMS
RESEARCH LABORATORY

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

September 1979

Army Project Number
2Q763743A780

Tactical Skill Acquisition
and Retention

Approved for public release; distribution unlimited.

IR00804

APPLICATION OF COMPUTER SIMULATION TECHNIQUES
IN MILITARY EXERCISE CONTROL SYSTEM DEVELOPMENT I:
NETMAN MODEL SENSITIVITY TEST AND VALIDATION

FOREWORD

Recent experience in military aviation and aerospace systems development has demonstrated computer modeling to be a well grounded and cost-effective design technique, especially for defining personnel requirements, i.e., manning levels, job descriptions, training, man-machine interfaces and operating procedures. Cognizant of this experience, the Army Research Institute for the Behavioral and Social Sciences (ARI) is conducting a research program investigating, in part, the application of computer simulation techniques in the design and development of Army field exercise management systems, from the standpoint of the man in the system.

The design of a field training and evaluation management control system focuses on field exercise management group responsibilities in staging unit training and/or evaluation exercises. Prominent in this research has been the development of a simulation model called NETMAN -- a stochastic digital computer model for simulating pertinent information throughput in a field exercise management system, with the emphasis on the people-portion of the process. Processing taxonomies simulated by NETMAN emerged from analyses of the Army Tactical Operations System (TOS), and, subsequently, the Marine Corps' Tactical Warfare Analysis and Evaluation System (TWAES).

The research reported here involving sensitivity testing, calibration and field validation of the NETMAN model was part of the larger research program designed to enhance field exercise management. As such, this research is part of Army Project 2Q763743A780, Training Development for Battlefield Effectiveness, and is responsive to the TRADOC Training Devices Directorate of the U.S. Army Training Support Center, Fort Eustis, Virginia.

JOSEPH ZEIDNER
Technical Director

EXECUTIVE SUMMARY

Problem

When large scale combat exercises are conducted, they are coordinated by a technologically advanced field exercise control system. To this end, correct and timely information must be obtained from the field by the control system in order to develop troop performance evaluations. Detailed information concerning the field situation is also necessary to control the insertion of scenario events such as artillery, aircraft maneuvers, enemy contacts, and other combat related situations.

The information in the system is also of value for deriving training requirements and for maintaining realism within the scenario. For example, casualties should coincide with the accuracy of an ordered artillery. In order for this realism to occur on the simulated battlefield, the field exercise control system must have current information from which to insert simulated casualties into the battle program. Accordingly, optimization of the exercise control system is important to achieving maximal benefit from costly field exercises.

The conceptual design of control systems is often carefully worked out. However, design test before system implementation is prohibitively expensive. Computer simulation represents a viable, relatively inexpensive alternative to field testing of such designs. A computer simulation model allows economical test of alternate design concepts. It is generally found that, although modification of design concepts as a result of computer simulation testing does not guarantee a perfect final system, it allows the elimination of many possibilities which will not work well in the field. It follows, then, that a reliable computer model which can be used to "try out" different exercise control system concepts and compare them in a quantified evaluation has strong appeal. Such a model could be used to help develop optimal field exercise control configurations and, thereby, maximize the training and troop evaluation benefits from field exercises.

Background

The Army Research Institute for the Behavioral and Social Sciences has developed a family of exercise control system oriented computer simulation models under contract with the Applied Psychological

Services. One model was developed to simulate message processing in the Army's Tactical Operations System (TOS). In the TOS system, messages composed by action officers are delivered to users of input-output devices for transmittal to a central computer and subsequent automatic updating of a battlefield information data bank. The computer model which simulates TOS is called MANMOD. MANMOD has been successively implemented and calibrated. Throughout its use, MANMOD has been found to be useful and reliable.

More recently, MANMOD was modified and expanded to allow simulation of an entire message processing network of the type used by a military field exercise control system. The resulting new computer model was called NETMAN. NETMAN was organized with the design of semiautomated military control systems specifically in mind. NETMAN assumes four levels of message processing. The message generation level is the exercise referee. Here, messages are generated for transmittal to the control center. The second level is message transmittal, in this case by the radio operator. The third level is a computer capable of decoding a message received from the field and presenting this message to the fourth level--a situation evaluator called a controller.

The NETMAN model is based on the prior highly tested MANMOD and has been restructured to accommodate message handling networks. Although some testing and model assessment was previously performed, further test, calibration, and validation analysis were necessary in order to assure confidence in the model in its present form.

Objectives

The overall objective of the current effort was to assess the current status of the NETMAN computer model. Determination was required of the degree of confidence which can be placed in the capability of the current configuration of the NETMAN model to evaluate exercise control systems. In addition to confidence assessment, an assessment of cost of use and the ease of use of the model was necessary. A final objective of the present work was to investigate and identify areas of NETMAN program modification so as to: (1) increase ease of use, (2) increase the fidelity of the model, (3) reduce the complexity of the input data preparation, (4) simplify the output so as to provide answers to specific questions, and (5) improve the utility of the model.

Methods and Results

A variety of computer runs was completed to test the performance of the NETMAN computer model under many different simulated conditions. The effect of these condition changes was then carefully evaluated from the viewpoints of rationality of output, cost analysis, ease of use, and program logic error.

The first NETMAN aspect investigated was output stability as a function of number of iterations. In a stochastic computer model like NETMAN, which can simulate numerous combinations of likely and unlikely events, a number of simulations of the same mission is required in order to arrive at a stable estimate of the output parameter. If an insufficient number of repetitions is used, the output will be sensitive to unlikely occurrences and may be biased. With more repetitions, however, the effect of unlikely events tends to balance out. The more complex the mission simulated, the larger the number of repetitions which are required to produce stable output. A relatively basic mission scenario was used in these tests. The results indicated considerable results stability with a limited number of iterations.

Parameters varied in the major sensitivity tests were: operator speed, operator precision, operator level of aspiration, operator stress threshold, operator fatigue level, number of operator networks, number of referee/radio operator teams per network, undetected error probability, message frequency, message length, transmission delay, and task difficulty. The results which were analyzed from such points of view as reasonableness, meaningfulness, utility, dependability, and reliability generally indicated support for the structural logic and internal validity of the model.

To determine the predictive validity of the NETMAN model, the Marine Corps' Tactical Warfare Simulation, Evaluation, and Analysis System (TWSEAS) was observed during control of a full battalion field exercise. Field observers, assigned to troop units, made measurements of message processing time and frequency. Additional observers collected data in the control center on message handling. In this manner, data were obtained concerning the quality of the operation of the TWSEAS system. Prior to the exercise, the control system personnel involved in the TWSEAS operation during the field exercise were tested to provide personnel operating characteristics for input to the model.

The NETMAN computer model was run to simulate the TWSEAS performance during the field exercise and the NETMAN-generated data were compared with the criterion data (actual TWSEAS operation). In

general, quite acceptable agreement was found between the model's predictions and actual TWSEAS operation. The predictions of the model fell, almost without exception, within one standard deviation of the exercise based data.

Implications

Due to the model response to the parametric variations of the sensitivity tests and the agreement between the model and the TWSEAS criterion data, a substantial degree of confidence may be placed in indications derived from the NETMAN computer model. Moreover, the model was implemented at a relatively low cost. Future exercise control system design would benefit from early test through the use of NETMAN. Moreover, the NETMAN model may be used to determine or confirm personnel allocation, effects of personnel proficiency, effects of various operator characteristics, network configuration, and the like in present network oriented exercise control systems.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
Simulation Models vs. Actual System Test	1
Overview of the NETMAN Model	2
Prior Message Processing Models	5
NETMAN Panel Review and Recommendations	6
Comments Concerning the Programming	7
Comments Concerning Mathematical Aspects	8
Comments Concerning General Utility	8
Suggestions Concerning the User Interface (Ease of Input and Output Interpretation)	9
Suggestions for Logic Changes	10
Panel Priority Weightings	10
Model Preparation for Sensitivity Tests	12
II. SENSITIVITY TESTS	15
Sensitivity Test Objectives	15
Simulation Results Rationality Confirmation	15
Cost to Run	15
Ease of Use	16
Program Error	16
Sensitivity Tests Performed	16
Variables and Levels	16
Number of Iterations	19
Results -- Rationality Confirmation	19
Operator Proficiency	20
Percentage Time Busy	24
Level of Aspiration	24
Stress Threshold	27
Fatigue	27
Number of Networks	31
Number of Teams per Network	36
Undetected Error Probability	39
Message Frequency	39
Field Message Length	43
Controller Message Length	45
Transmission Delay	48
Task Difficulty	48
Cost to Run	50

TABLE OF CONTENTS (cont.)

	Page
Ease of Use	53
Program Error	54
III. MODEL VALIDATION	55
Criterion Data Attributes	55
Criterion Variables Selected	55
Model Validation Concepts	59
Model Validation	64
Individual Parameter Data Development	65
Task Performance Time for Referees	65
Task Performance Time for Radio Operators	70
Stress Threshold Determination	72
Level of Aspiration Determination	74
Individual Precision Determination	74
Individual Speed Factor	76
Background Information	79
Message Frequency and Frequency of Message Type	79
Computer Runs Performed	81
Observer Training	81
Field Exercise	83
Field Exercise Data Collection	83
Discussion of Field Exercise Data	85
Validation Analysis	86
Thoroughness	86
Responsiveness	86
Overall Effectiveness	87
Total Message Processing Time	88
Operator Processing Time	88
Transmission Delay	89
Integration of Validation Data	89
Discussion	90
IV. SUMMARY AND CONCLUSIONS	95
REFERENCES	99

LIST OF TABLES

Table		Page
1-1	Panel Priority Rankings of Major Issues	11
2-1	Sensitivity Tests Parameters	17
2-2	Summary of Variance Analyses of Operator Speed and Precision	21
2-3	Percentage of Time Busy as a Function of Operator Speed and Precision	26
2-4	Effects of Stress Threshold on Simulation Results	28
2-5	Effects of Fatigue Variable on Mean Times per Message and on Thoroughness	30
2-6	Summary of Analyses of Variance of Fatigue Effects	32
2-7	Effects of Number of Networks Parameters on Simulation Results	33
2-8	Summary of Analyses of Variance for Network Size Effect on Message Completion	35
2-9	Summary of Analysis of Variance for Network Size Effect on Time	35
2-10	Effect of Number of Teams Per Network on Effectiveness Measures	38
2-11	Summary of Analysis of Variance for Undetected Error Probability Effect	38
2-12	Summary of Analysis of Variance of Message Frequency Effect on Overall Effectiveness	42
2-13	Summary of Analysis of Variance for Task Difficulty Effect on Messages Processed Per Hour	42
2-14	Summary of Analysis of Variance for Task Difficulty Effect on Message Processing Time	52
2-15	Summary of Analysis of Variance for Task Difficulty Effect on Overall Effectiveness	52

LIST OF TABLES (cont.)

Table		Page
3-1	Desirable Criterion Attributes	56
3-2	Criterion Variable Description and Method of Measurement	57
3-3	Criterion Evaluation	58
3-4	Criteria for Evaluating the Utility of a Computer Model	60
3-5	Referee Task Analyses	69
3-6	Radio Operator Task Analytic Time	73
3-7	Stress Threshold of Each Referee and Radio Operator	73
3-8	Level of Aspiration of Each Referee and Radio Operator	75
3-9	Success Probability and Precision of Each Referee and Radio Operator	77
3-10	Mean Messages Processing Time and Speed Factor for Each Referee and Radio Operator	78
3-11	Field Message Frequency per Hour	80
3-12	Message Type Frequency	80
3-13	Consolidation of Message Types	82
3-14	Summary of Validation Statistics	92

LIST OF FIGURES

Figure		Page
1-1	Schematic of message processing flow	3
2-1	Operator level of proficiency predictions	22
2-2	Mean time per message as a function of referee (a), radio operator (b), and controller (c) speed and precision	23
2-3	System responsiveness (a) and thoroughness (b) as a function of simulated personnel speed and precision	25
2-4	Messages completed as a function of number of teams per network	34
2-5	Effect of number of teams on time spent by controller on message processing	37
2-6	Effect of undetected error probability on the accuracy component of effectiveness	40
2-7	Effect of number of messages per hour on overall effectiveness and on effectiveness components	41
2-8	Effect of message length on messages completed and mean time per message for referee and radio operator	44
2-9	Effect of message length on messages completed and mean time per message for controller	46
2-10	Effect of transmission delay on processing time	47
2-11	Effect of task element success probability on message processing times	49
2-12	Effect of task element success probability on effectiveness elements and on overall effectiveness	51
3-1	Steps in model development	62
3-2	Information provided to referees for composing one TWSEAS message	66

LIST OF FIGURES (cont.)

Figure		Page
3-3	Task analytic data collection form for referees	67
3-4	Task analytic data collection form for radio operators . . .	71
3-5	Field data collection form	84
3-6	Model predictions versus field data	91

I. INTRODUCTION

The NETMAN model was developed to provide a basis for stochastic simulation of information throughput in a military field exercise control system. This model, described briefly later in this chapter, permits its users to evaluate such effects as personnel distribution, varying system configurations, training, and workload on system performance in tactical military exercise control systems.

The NETMAN model is a second generation message-handling simulation model. Its predecessors are summarized later in this chapter. Although the prior work provided preliminary sensivity testing of NETMAN, a more thorough program of sensitivity testing and validation was determined to be required.

Simulation Models vs. Actual System Test

Inherent in any computer simulation concept is the understanding that considerable savings can be achieved by substituting computer simulated exercises for actual system test. Such savings, of course, are predicated on a demonstration that adequate agreement can be achieved between a model, such as NETMAN, and actual system operation. A second advantage of the use of a model is that the relative time required to yield visable results is less than that of an actual system exercise. The initial data collection and preparation for a model simulation may be extensive but, once complete, the results of parametric variations may be obtained in very short times (minutes/hours) as is shown in the sensitivity runs of Chapter 2. Other advantages of models in general over actual system exercises are:

- exercising a model is less costly
- fewer personnel are involved
- models are independent of uncontrollable conditions
- models do not expose personnel to danger or accidents
- models do not expose equipment to damage
- convenience

Note also that the more complex, costly or large scale the operational system, the more dominent these relative advantages of simulation modeling become.

Last, simulation offers the capability to consider and evaluate the impact of new anticipated equipment, different speeds and numbers of communication lines, as yet unauthorized operator sequences/procedures, and system loads.

Overview of the NETMAN Model

The NETMAN model simulates each person and each message involved in the data acquisition required for evaluating performance during field exercises. These personnel include up to 27 referees, 27 radio operators, and 3 controllers interacting in a fixed network of communication lines linked with a Central Computing Center (CCC).

The field exercise data acquisition situation simulated may be viewed as a message processing network configured as shown in Figure 1-1. This figure symbolically displays 27 simulated referees (R_1 through R_{27}) receiving simulated symbolic input messages from independent sources as well as from one of three simulated controllers (CON 1, CON 2, or CON 3). Messages are indicated by the symbol ~.

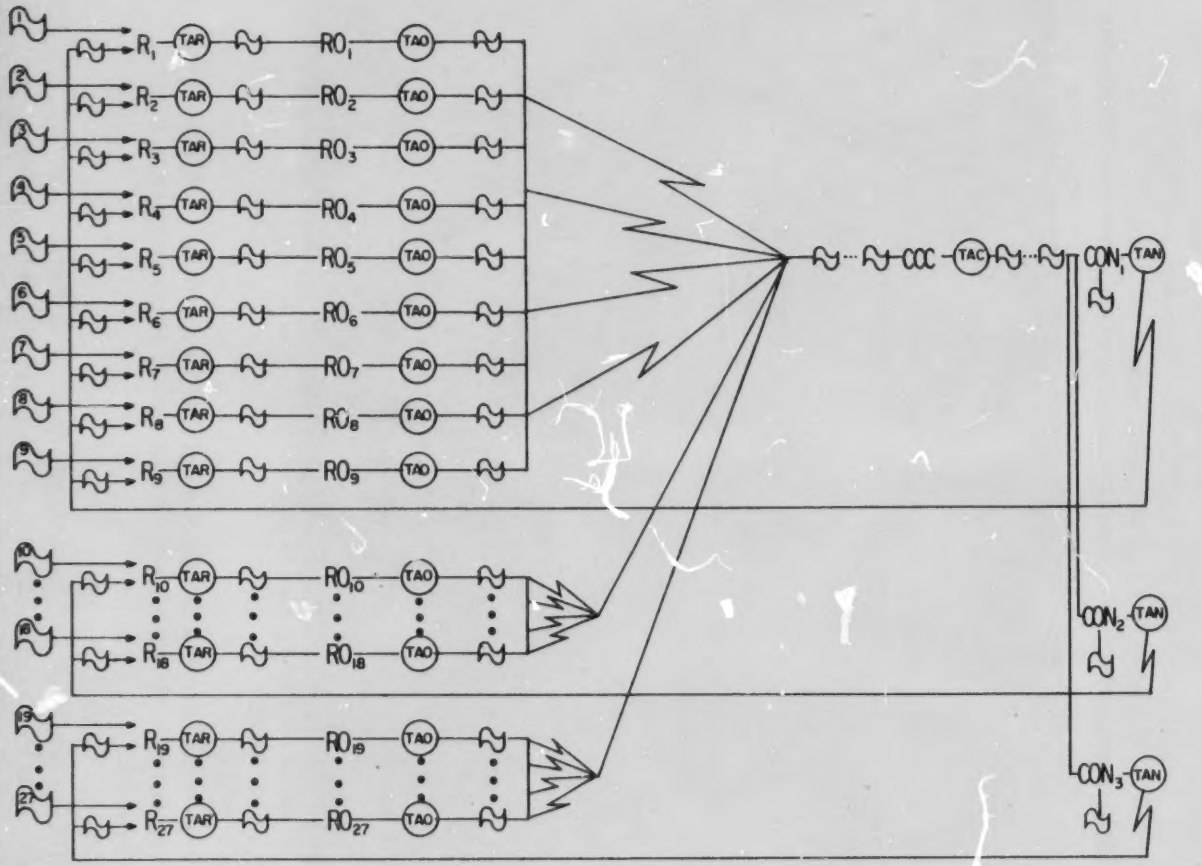
Military field exercises of some forms are observed by referees, who complete evaluative and situational reports which are transmitted to a computer via a radio operator. Messages introduced into the system are processed by various personnel and the CCC and then delivered to controllers for evaluation on CRT terminals. The field exercise data acquisition is simulated through random message generation based on pertinent values such as message length, type, and arrival time. Each generated message is then processed through the referee → radio operator → CCC → controller network and processing time is determined along with a number of other descriptive indices.

Each simulated referee in the left to right message processing flow of Figure 1-1 performs some appropriate procedure on the simulated message(s) received. This is shown symbolically by a circle in which TAR (Task Analysis, Referee) appears. A message then passes to the corresponding radio operator (one of RO_1 through RO_{27}) for processing in accordance with some specified task analysis procedure, circle TAO.

Up to 27 simulated messages, each processed by a different radio operator, could then be ready for entry into the CCC. Entry into the simulated CCC for any given message is made over the communication line for the three networks shown. Accordingly, in a given network, there may be up to nine simulated messages competing for the one available CCC input line.

Telecommunication lines are designated in Figure 1-1, and the resultant queues awaiting CCC actions are designated by ~ ··· ~.

On a first-in, first-out basis, the CCC processes messages from all of the three networks in accordance with its task analysis procedure--depicted by circle TAC (Task Analysis, Computer). These messages then enter another queue awaiting action by one of three controllers. Each



TAR = TASK ANALYSIS, REFEREE; TAO = TASK ANALYSIS, RADIO OPERATOR; TAC = TASK ANALYSIS, COMPUTER
TAN = TASK ANALYSIS, CONTROLLER

Figure 1-1. Schematic of message processing flow.

simulated controller then assesses and operates on the oldest message from his network in the queue and performs in accordance with the task analysis for controllers as symbolized by circle TAN.

The loop is closed by the simulated generation of new messages by the controller for input to one of the referees in his nine-way network as a function of a parameter input to the model. Besides the link from the controller to his referees, the referees are also interconnected and may generate new messages as a result of their interactions with one another. Up to 5,400 messages may be included in a given simulation run. During NETMAN's processing of message information, the model places special emphasis on certain human performance features considered to be important in a field operational system of this sort. These include operator stress, fatigue, and level of aspiration.

The overall man-machine performance measure calculated by the model, effectiveness, is calculated for each day of the simulated exercise. It is composed of four independent factors--thoroughness, completeness, accuracy, and responsiveness.

Output from the model is presented in the form of computer tabulations and terminal displays. These provide data which promote insight for evaluating alternatives both in terms of absolute and relative value. The printed and displayed output is designed and organized so as to provide results that answer questions of practical importance, such as:

1. What is the average processing time for a message through the network as a function of the frequency of input messages, operator capabilities, and network configuration?

2. How do changes of input parameters affect predicted total man-machine effectiveness values?

3. What is the loading situation relative to idle vs. busy time for referees, radio operators, the CCC, and the controller?

4. How great a stress is placed on the operators? What is the fatigue profile over the mission for each type of personnel?

5. Would changes in the task organization of one or more operators materially affect average processing time and system effectiveness?

6. What are some effects of operator commission and omission errors under various conditions?

7. How would increased personnel training or improved personnel selection affect system performance?

The program presents detailed message processing time and error information, if desired, as well as hourly summary and run summary outputs. The detailed message processing output shows the fine grain of the results of the simulation of each task in the processing of messages.

The hourly summary presents a consolidation of the results of a simulated hour's work across all iterations and includes items such as: number of messages completed, time spent working, end of hour stress level, performance and aspiration, time spent performing various processes, and average time per message.

The simulation run summary, produced after N iterations of the exercise, includes manpower utilization, message processing times, overall effectiveness indicators, and workload summary information.

NETMAN is programmed in FORTRAN IV for the Univac 1108 system. It is organized to allow the user to conduct various numerical experiments relative to the field exercises. Each computer run of the model represents a simulation of a field session up to 10 hours in duration conducted under conditions as specified by input parameters. Examples of exercise input parameters include the frequency of messages entered to the system, the number of operators, and the speed and aspiration levels of these operators.

A model description is contained in Siegel, Leahy and Wolf (1977) together with a discussion on model utilization, program flow charts, subroutine definitions, user input-output formats, and task analyses.

Prior Message Processing Models

Several of the NETMAN model concepts are based on a prior operational computer model, developed by Applied Psychological Services in collaboration with the U.S. Army Research Institute for the Behavioral and Social Sciences, for simulating the U.S. Army's Tactical Operations System (TOS). The earlier model, called MANMOD (Siegel, Wolf, & Leahy, 1972) simulates the behavior and performance of up to six men who function as action officers and input-output device operators in the TOS system. These men perform tasks similar to those simulated in NETMAN. The mechanism for task-by-task performance evaluation in NETMAN is basically the same as in MANMOD.

MANMOD was originally designed for batch run processing in FORTRAN IV on the CDC 3300 computer. In this form, original sensitivity

and validation runs were made. In the validation, a high degree of correspondence was found between the model's output and a set of error data collected from an independent source.

In a follow-on effort (Siegel, Wolf, Leahy, Bearde, and Baker, 1973) the MANMOD model was modified to operate in an interactive computer time-sharing mode. This feature allows the experimenter (mission analyst) to interact in a "conversational" mode with the model and to enter data "on line." This interaction is performed through a computer terminal and greatly increases the ease with which simulations can be performed. NETMAN possesses this same type of interactive capability.

A variant of the MANMOD was also developed (Leahy, Lautman, Bearde, and Siegel, 1974) which allows collection of data during an experiment in which one or more actual operators perform a part of the process and the computer simulates the remainder of the TOS activity.

More recently, MANMOD was adapted for the Univac 1108 computer and several new capabilities were added which increase the realism of the simulation. It was modified to exchange data with two other independent computer models in such a way as to maximize the strong points of each of the models (Leahy, Siegel, and Wolf, 1975a, 1975b). Other areas of similarity between MANMOD and NETMAN are:

1. the message generation technique is similar

2. the operator performance and aspiration determination technique is the same

3. some of the parameters and two of the four effectiveness factors are the same

4. the basic nature of both models is stochastic. As a result, a number of repetitions is required to produce a stable result

5. both models provide lists of inputs, optional detailed output, hourly summaries, and run summaries.

NETMAN Panel Review and Recommendations

Before proceding with the NETMAN sensitivity tests and validation, it seemed proper to obtain an independent review of the model. A review panel of independent personnel, expert in various areas of digital modeling and related fields, was brought together for this purpose. The panel members made suggestions relative to: model improvement, additional features to be added, features which should be deleted, features which may yield misimpressions, errors in the work, and attributes which are useful. Particularly important were suggestions relative to improvements concerning documentation, sensitivity test considerations, and validation.

In preparation for the panel meeting, panelists were advised of the following areas of emphasis:

- How the programming, its architecture and organization, the documentation, and the programming structure might be improved.

- How the mathematics could be improved, any errors, any discontinuities, any area over embellished, any "better" approaches.

- What is the general utility of the model for achieving its purpose, how the model compares with other models, how the documentation compares with documentation of physical models and obvious positive and negative aspects.

- The value of the user-model interface, the value of the model, and the output formats as aids in decision making along with display characteristic, output interpretation, and system architecture improvements.

- How the Army user will use such a model, where he will find difficulty with it, where he will find its information useful, what can be done to improve its utility, what are the user interests and characteristics, and what input problems face the user.

A summary of the panel's conclusions in each major area follows.

Comments Concerning the Programming

The panel found the model to be highly portable, being written in FORTRAN IV with almost no machine specific aspects. The exception is the random number generator. Such routines are usually uniquely developed for each computer system.

The program was held to follow the rules of structured programming. Each major function in the program is separated into its own subroutine. This allows program changes to be quickly made.

The subroutines were said to be well organized and the interfaces between subroutines to be clearly identified. Internal documentation, including comment cards, was held to be adequate.

The fixed format input was said to be difficult for the user and substitution of free format input or more identifying information was

suggested. The panel also said that the program is relatively large. This feature may limit the number and type of model enhancements. Additionally, the panel contended that some data are specified for input but not used internally (for example, operator factors for the computer are not used).

A minor discrepancy in the program was noted by the panel. This discrepancy appears to have no effect on the processing: If the random number generator is called in a certain way, an erroneous output would result. However, the random number generator is never called in this way. Accordingly, a problem could only arise in the case of model changes which involved calling the random number generator in the specific error application.

Comments Concerning Mathematical Aspects

The stress function has a strong discontinuity at 95 percent of time worked. This may cause an unrealistic instability in simulated performance when the percentage time worked is around 95%. It was noted that only part of the originally developed stress function was incorporated into NETMAN and that inclusion of the entire stress function should be considered.

The message length variable is simulated in the model with a normal distribution, while the real distribution would probably be positively skewed. It was suggested that a log normal distribution or a triangular distribution might be substituted.

Comments Concerning General Utility

The panel held the model to be flexible and to be useful for simulating exercise control systems. Because the model is totally compartmentalized, it can be extended to other goals not originally anticipated.

NETMAN is a research tool and as such is primarily intended for research psychologists and system analysts. It offers a way of trying out new ideas, procedures, equipment configurations, and personnel combinations. Generally, such models have value in that "if it works in the model, it may work in the field. If it does not work in the model, then it will not work in the field." NETMAN is not suitable for a casual user and needs more user choice of output if it is to be directly useful for system engineers.

The input is a formidable obstacle for first-time users, according to the panel. The values to assign to the input data are also a problem. There are little data available from field exercises, and laboratory data have limitations.

Suggestions Concerning the User Interface (Ease of Input and Output Interpretation)

The panel's recommendations for the user interface included:

1. User worksheets should be developed which would lead a user through the input data collection and preparation in a step-by-step manner.

2. Examples of how to set up input data should be included in the User's Manual.

3. Automatic edits of input data to identify illegal or erroneous entries should be provided.

4. More data default options are needed to facilitate the case in which the user only wants to enter a small set of new data.

5. A capability is needed which would allow the user to enter data in the form that is normally available and have the computer automatically perform any transformations required. For example, the program now requires the probabilities of related events to be entered as cumulative probabilities.

6. An option to print as results only the data type which has been changed, instead of all of a category, is needed.

7. In the interactive mode, the need exists for more computer prompting and options for data display.

8. A more positive action is required to initiate execution. At present, a blank or zero entry starts the processing.

9. The operator should be able to display only selected output data elements or only the data elements that change between simulations.

10. To facilitate understanding, graphic output displays should be provided.

11. An easy interface to statistical program packages which perform data analysis might be helpful.

Suggestions for Logic Changes

The panel's suggestions for logic changes included the need for entering human parameters for each man individually (at present, these parameters are entered as constant for each type of personnel) and for the inclusion of stress, aspiration, and fatigue in the error generation formula. The panel also suggested different fatigue curves for different personnel types since it felt that the referees and the radio operators in the field may fatigue more quickly than controllers.

The panel also suggested that logic might be added to simulate referee mobility, simultaneous observation of multiple situations, and decision making functions. At present the referee is primarily simulated only in his message originating aspect.

Fluctuation of level of aspiration due to local effects such as presence of officers and situational variables was indicated by the panel to be an additional desirable feature along with an increase in the simulation duration capability beyond 10 hours.

Panel Priority Weightings

A list was compiled of the 22 major issues raised during the panel discussion and each of the six panel participants was asked to judge each item on the following five point scale:

5 = High priority
4 = Considerable priority
3 = Moderate priority
2 = Low priority
1 = Negligible priority

After priorities were assigned by each panel expert to each item, the mean priority assigned was calculated for each item. The mean priority ratings were ranked. The items, as well as the ranking of the priorities, are shown in Table 1-1. The three items tied for the highest priority were user manual enhancements, programming changes for improved ease of use, and discrepancies between manual and program. The fourth ranked item was utility testing, which refers to use of the model in such a way as to draw out any problems, limitations, or other unknown characteristics of the program. This item is largely satisfied by the sensitivity tests and validation runs, reported in later chapters of this report.

Tied for fifth place were norm development, data base development, and abbreviated user mode initialization. Norm development refers to the availability of the necessary input data for the model, and abbreviated user mode includes the reduction of data input and output by the user to a level most convenient for his application.

Table 1-1

Panel Priority Rankings of Major Issues

Issue No.	Description of Issue	Issue Rank
1.	User manual enhancement	2
2.	Programming changes for improved ease of use	2
3.	Discrepancies between manual and program	2
4.	Utility testing	4
5.	Norm development, data base development	5
6.	Abbreviated user mode, initialization	5
7.	Human effects on error rate incorporation	8
8.	Inquiry mode for output and output difference analysis	8
9.	Overall processing	8
10.	Better simulation of referee activity	10
11.	Graphic output provision	11
12.	Window package display: controller-referee interface	11
13.	Effectiveness measures	13
14.	Fatigue curve individualization	13
15.	Stress function-error frequency modification	17
16.	Catastrophic failure indicator	17
17.	Sensitivity to local stressors	17
18.	Statistical package incorporation	17
19.	Human submodel validation	17
20.	Validity	20
21.	Geographic representation	21
22.	Stress threshold test	22

Tied for eighth place were human effects on error rate, inquiry mode for output, and overall processing. The human effects which might be expected to affect error generation are stress, aspiration, and fatigue. The inquiry mode would allow an on-line experimenter to request the specific sets of data or analyses which are relevant for his use. Overall processing refers to the entire throughput and the event oriented sequencing of message handling.

Tenth ranked was the expanded simulation of referee activity described earlier under logic changes.

Lower priority issues are considered next. These include: (a) graphic output to allow the capability of automatically transforming the model output into bar graphs or line graphs, (b) a window package to allow a special type of output display, (c) fatigue curve individualization to allow different decrement curves for different levels of the network as well as different decrements for different operations at the same level of the network, (d) stress function error frequency modification to allow stress to affect the number of errors, (e) catastrophic failure indication to allow consideration of those factors, such as radio failure, which would actually shut down or greatly reduce the effectiveness of the system, (f) sensitivity increase as the result of local stressors such as level of aspiration change in the presence of officers, (g) statistical package interface so as to facilitate data reduction, (h) separate testing, perhaps through laboratory measurement, of the component subroutines which simulate human activities, (i) incorporation of area and movement effects on personnel in the simulation, and (j) investigation and possible incorporation of a fuller stress function in the NETMAN model.

Model Preparation for Sensitivity Tests

A full description of the NETMAN program is found in Leahy, Siegel, and Wolf (1975). To facilitate the sensitivity and validation runs, a series of fine tuning and readiness-type adjustments was made to the program. These are summarized here:

1. the prior NETMAN program was copied to a new file (NETVAL) and an intact copy of NETMAN was retained so as to allow separation of the two developments.

2. new control language files were prepared to facilitate simulations.

3. the input format was expanded to allow entry of the day number for fatigue calculation for each simulated person rather than a single value for all individuals.

4. the input format was changed to allow input of different values of speed, precision, aspiration, and stress threshold for each person simulated.

5. message origin identification was revised so that messages originating from the same stimulus message are keyed to that source message.

6. a new type of task element, random walk task, was added to better simulate decision tasks. Associated with this a new subroutine, INRAN, was added to control random walk data input.

7. corrections were made in the random number generation function to prevent its misuse as discussed in the preceding section.

8. variable dimension statements were modified to accommodate larger data arrays for the following: fatigue, speed, precision, aspiration, and stress threshold.

9. modification to allow simulation of networks which include one referee, one radio operator, and one controller--a case not previously handled by NETMAN.

10. set up of new files for input, output, and mapping, compiling, and printing elements. These changes facilitate performance of multiple simulation runs.

11. enhancements to enable the analyst to enter task difficulty and task duration data from the terminal in an interactive mode.

II. SENSITIVITY TESTS

As the first step in evaluating the utility of the NETMAN simulation model, a comprehensive set of sensitivity tests was defined and conducted. The purpose of these tests was to evaluate the internal logic of the model, as a prerequisite to a meaningful validy study, presented in Chapter III.

Sensitivity Test Objectives

Four inherent aspects of the NETMAN program were considered in the sensitivity tests: simulation results rationality confirmation, cost analysis, ease of use, and logic error.

Simulation Results Rationality Confirmation

One of the major goals of the sensitivity test of any model is confirmation that the model's output is sensitive to variation in model input data. Not only should the output reflect changes in input but the change should possess appropriate directionality and magnitude. Moreover, the output should, in most cases, be reasonably smooth over a change in input values. Parametric values were selected for the sensitivity tests which would allow determining whether or not the model exhibits these properties.

Sensitivity test results often suggest a need for model changes. Within a model, constants and equations are employed to detail relationships. When an illogical result is obtained, the need for modifying such constants and relationships is suggested.

In addition to errors of commission, errors of omission in the program were sought as the result of sensitivity tests. There may be combinations of conditions for which additional logic is required or the need for additional variables may be suggested. When developing a model, it is difficult to anticipate all possible interactions. Sensitivity tests were made with the thought that the model would be calibrated as required.

Cost to Run

A byproduct of sensitivity tests is information concerning the cost for running the program. This was a secondary goal of the tests. Although the cost of computer time is generally slight in comparison with the cost of other methods of simulation, computer time cost may be a limiting factor in model use. Knowledge about run costs as a function of input parameters was therefore sought.

Ease of Use

Information about the ease of use of the program is another by-product of the process of organizing and performing sensitivity tests. Problems relative to input data organization and program execution become evident. Additionally, estimates of the time required to prepare the input data are made available.

Program Error

The NETMAN program was previously tested and is believed to contain minimum logic error. However, one goal of the sensitivity testing is identification of any overlooked program errors. Program error correction is an ongoing process which starts in the initial programming, continues on through the sensitivity testing, and often continues even into final product use, with the return from tests becoming smaller and smaller as additional tests are performed. Sensitivity tests allow examination of the model in many diverse situations and elimination of program errors as they are identified.

Sensitivity Tests Performed

In order to meet the sensitivity test objectives described above, a comprehensive set of sensitivity tests was performed. The parameters varied in these tests are shown in Table 2-1. The program was run with the parameters shown on a Univac 1108 Computer System. Both the terminal and the batch modes of operation were employed.

Variables and Levels

The tests included a large number of variables, and each variable was tested at two or more levels including the limits as specified in the model. Generally, when a single input parameter was varied across its practical range, all of the other variables in the model were held constant. In addition to the parametric variation, a number of selected combinations of variables was tested.

The primary input parameters were selected for variation in the sensitivity tests:

1. number of iterations of the simulation run
2. operator speed measure } operator proficiency
3. operator precision level }
4. operator aspiration level
5. operator stress level
6. operator fatigue level
7. number of operator networks
8. number of operator teams per network

Table 2-1

Sensitivity Test Parameters

Run Number	Parameter Varied	Assigned Values: Iterations	Operator speed	Operator precision	Operator level of aspiration	Operator stress threshold	Days of work (fatigue)	Number of networks	Number of teams per network	Undetected error probability	Message frequency per hour per referee	Field message length	Controller message length	Transmission delay	Task difficulty	Task duration	Mission duration (hours)
1a	Iterations	5	1.0	1.0	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1x		N	1.0	1.0	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
2	Operator Proficiency (speed and precision)	N	.90	.90	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
3		N	.95	.90	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
4		N	1.00	.90	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
5		N	1.05	.90	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
6		N	1.10	.90	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
7		N	.90	.95	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
8		N	.95	.95	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
9		N	1.00	.95	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
10		N	1.05	.95	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
11		N	1.10	.95	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
12		N	.90	1.00	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
13		N	.95	1.00	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
14		N	1.00	1.00	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
15		N	1.05	1.00	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
16		N	1.10	1.00	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
17		N	.90	1.05	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
18		N	.95	1.05	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
19		N	1.00	1.05	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
20		N	1.05	1.05	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
21		N	1.10	1.05	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
22		N	.90	1.10	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
23		N	.95	1.10	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
24		N	1.00	1.10	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
25		N	1.05	1.10	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
26		N	1.10	1.10	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
27	Operator Aspiration	N	1.00	1.00	.97	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
28		N	1.00	1.00	.90	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
29		N	1.00	1.00	.98	2.3	1	2	8	.01	10	22	300	2.0	.95	2	4
30		N	1.00	1.00	.95	2.3	1	2	8	.01	10	22	300	2.0	.95	2	4
															.95		
31	Operator Stress Threshold	N	1.00	1.00	.95	2.0	1	2	8	.01	5	22	300	2.0	.95	2	4
32		N	1.00	1.00	.95	2.0	1	2	8	.01	15	22	300	2.0	.95	2	4
33		N	1.00	1.00	.95	3.0	1	2	8	.01	5	22	300	2.0	.95	2	4
34		N	1.00	1.00	.95	3.0	1	2	8	.01	15	22	300	2.0	.95	2	4

Table 2-1 Continued

35	Fatigue	N	1.00	1.00	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	10
36		N	1.00	1.00	.95	2.3	5	2	8	.01	5	22	300	2.0	.95	2	10
37		N	1.00	1.00	.95	2.3	9	2	8	.01	5	22	300	2.0	.95	2	10
38	Number of	N	1.00	1.00	.95	2.3	1	1	8	.01	10	22	300	2.0	.95	2	4
39	Networks	N	1.00	1.00	.95	2.3	1	2	8	.01	10	22	300	2.0	.95	2	4
40		N	1.00	1.00	.95	2.3	1	3	8	.01	10	22	300	2.0	.95	2	4
41	Number of	N	1.00	1.00	.95	2.3	1	2	1	.01	5	22	300	2.0	.95	2	4
42	Teams per	N	1.00	1.00	.95	2.3	1	2	5	.01	5	22	300	2.0	.95	2	4
43	Network	N	1.00	1.00	.95	2.3	1	2	9	.01	5	22	300	2.0	.95	2	4
44	Undetected	N	1.00	1.00	.95	2.3	1	2	8	.1	5	22	300	2.0	.95	2	4
45	Error	N	1.00	1.00	.95	2.3	1	2	8	1.0	5	22	300	2.0	.95	2	4
46	Message	N	1.00	1.00	.95	2.3	1	2	8	.01	15	22	300	2.0	.95	2	4
47	Frequency	N	1.00	1.00	.95	2.3	1	2	8	.01	25	22	300	2.0	.95	2	4
47A		N	1.00	1.00	.95	2.3	1	2	8	.01	30	22	300	2.0	.95	2	4
48	Field	N	1.00	1.00	.95	2.3	1	2	8	.01	5	10	300	2.0	.95	2	4
49	Message	N	1.00	1.00	.95	2.3	1	2	8	.01	5	100	300	2.0	.95	2	4
50	Length	N	1.00	1.00	.95	2.3	1	2	8	.01	5	1000	300	2.0	.95	2	4
51	Controller	N	1.00	1.00	.95	2.3	1	2	8	.01	5	22	50	2.0	.95	2	4
52	Message	N	1.00	1.00	.95	2.3	1	2	8	.01	5	22	100	2.0	.95	2	4
53	Length	N	1.00	1.00	.95	2.3	1	2	8	.01	5	22	300	2.0	.95	2	4
54	Transmis-	N	1.00	1.00	.95	2.3	1	2	8	.01	5	22	300	0	.95	2	4
55	sion	N	1.00	1.00	.95	2.3	1	2	8	.01	5	22	300	10.0	.95	2	4
56	Delay	N	1.00	1.00	.95	2.3	1	2	8	.01	5	22	300	100.0	.95	2	4
57	Task	N	1.00	1.00	.95	2.3	1	2	8	.01	5	22	300	2.0	.99	2	4
58	Difficulty	N	1.00	1.00	.95	2.3	1	2	8	.01	5	22	300	2.0	.80	2	4
59	(Success Prob.)	N	1.00	1.00	.95	2.3	1	2	8	.01	5	22	300	2.0	.60	2	4

9. undetected error probability
10. message frequency (number per hour)
11. message length (characters)
12. controller message length (characters)
13. duration of transmission delay (seconds)
14. task difficulty (task success probability)

Number of Iterations

One major consideration in the cost of employing a stochastic model is the number of simulations or iterations required to produce a stable estimate of output parameters. If the sample is too small, chance combinations of low probability events may bias the resulting simulation summaries. In order to avoid this random error a sufficiently large number of iterations must be completed. The minimum size of this sample is a function of the variation in the various random processes embedded in the simulation.

The data used to evaluate stability of simulation outputs were the mission segment times which identify completion points at various stages of message processing. There are 19 segment times given for each of seven message types. The resulting 133 time values provide a sufficient N to make stability comparisons between computer runs, i.e., between sets of computer iterations of the simulated exercise. Pairs of simulations were run with increasing numbers of iterations. The results of the simulations were intercorrelated and the mean difference between runs was calculated. The criteria for sufficient iterations was the point at which the correlation between simulations was .90 or greater and /or the mean difference between runs was five percent or less.

The Pearson product moment correlation was calculated between the segment time data for each mission segment from two independent five-iteration simulations. Each such run was initiated with identical input parameters except for different starting pseudo-random numbers. The resultant correlation coefficient was .9999. This indicates a very high degree of stability for the simulation involved. Such high stability might not be attained for more complex scenarios which involve more branching. The mean difference between runs was less than five percent.

Results -- Rationality Confirmation

The major result of the sensitivity tests was confirmation that the model produces results which vary with input variation in a reasonable manner. The results of various simulation runs which were completed relative to this issue are presented below.

Operator Proficiency

Operator proficiency is reflected in the NETMAN model through two variables--operator precision and speed. Operator precision affects the probability of error while operator speed affects the amount of time required to perform tasks. These variables were combined factorially to allow analysis of the main effects as well as the interactions. The value of each of these two parameters was varied as follows: 0.90, 0.95, 1.00, 1.05, and 1.10. These runs are listed in Table 2-1 as runs 2 through 26. Other exploratory runs were made with a setting of 0.5 for operator speed. Three separate two-way analyses of variance were performed on the segment time data from this set of 25 runs, for referee, radio operator and controller. The results of these three variance analyses are shown together in Table 2-2. The results of these analyses showed that the following parameters produced statistically significant effects:

speed of referee
speed of radio operator
speed of controller

precision of referee
precision of radio operator
precision of controller

interaction of speed and precision for referee
interaction of speed and precision for controller

The anticipated direction of each prediction is indicated in Figure 2-1 by an arrow head. Prediction number 1, as shown by this Figure is that simulated operators with a precision of .90 and a speed of .90 will perform better than simulated operators with a precision of .90 and a speed of .95. Similarly, prediction 27 is that simulated operators with a speed of .95 and a precision of .95 will do better than operators with a speed of .95 and a precision of 1.00.

There were 120 predicted differences (40 each for referee, radio operator, and controller). Of these, 113 were in the correct direction, six were tied and one was in the wrong direction. A normal curve approximation yielded a statistical significance level of 0.001 for this result.

Figure 2-2 (a), (b), and (c) show the mean time in seconds required per message for each of the three types of system personnel as a result of the variation of the operator speed and precision parameters over the range 0.90 to 1.1. As expected, in all cases, improved speed and precision resulted in reduced message processing time. For speed, the value 0.9 represents fast persons and 1.1 represents slow personnel. Precision refers to the operator tendancy to produce errors which require repeat of

Table 2-2

Summary of Variance Analyses of Operator Speed and Precision

Operator Position	Source	Degrees of Freedom	Sum of Squares	Mean Square	F
Referee	Total	99	55,991.39		
	Speed (A)	4	5,056.94	1,264.24	151.59**
	Precision (B)	4	50,047.54	12,511.89	1,500.23**
	A x B	16	261.16	16.32	1.96*
	Error	75	625.75	8.34	
Radio Operator	Total	99	81,764.75		
	Speed (A)	4	5,405.60	1,351.40	96.39**
	Precision (B)	4	74,935.10	18,733.78	1,336.22**
	A x B	16	372.30	23.30	1.66 N.S.
	Error	75	1,051.25	14.02	
Controller	Total	99	52,958.51		
	Speed (A)	4	5,111.06	1,277.77	148.92**
	Precision (B)	4	46,907.56	11,726.89	1,366.77**
	A x B	16	296.64	18.56	2.16*
	Error	75	643.25	8.58	

* $p \leq .05$

** $p \leq .001$

N.S. Non significant

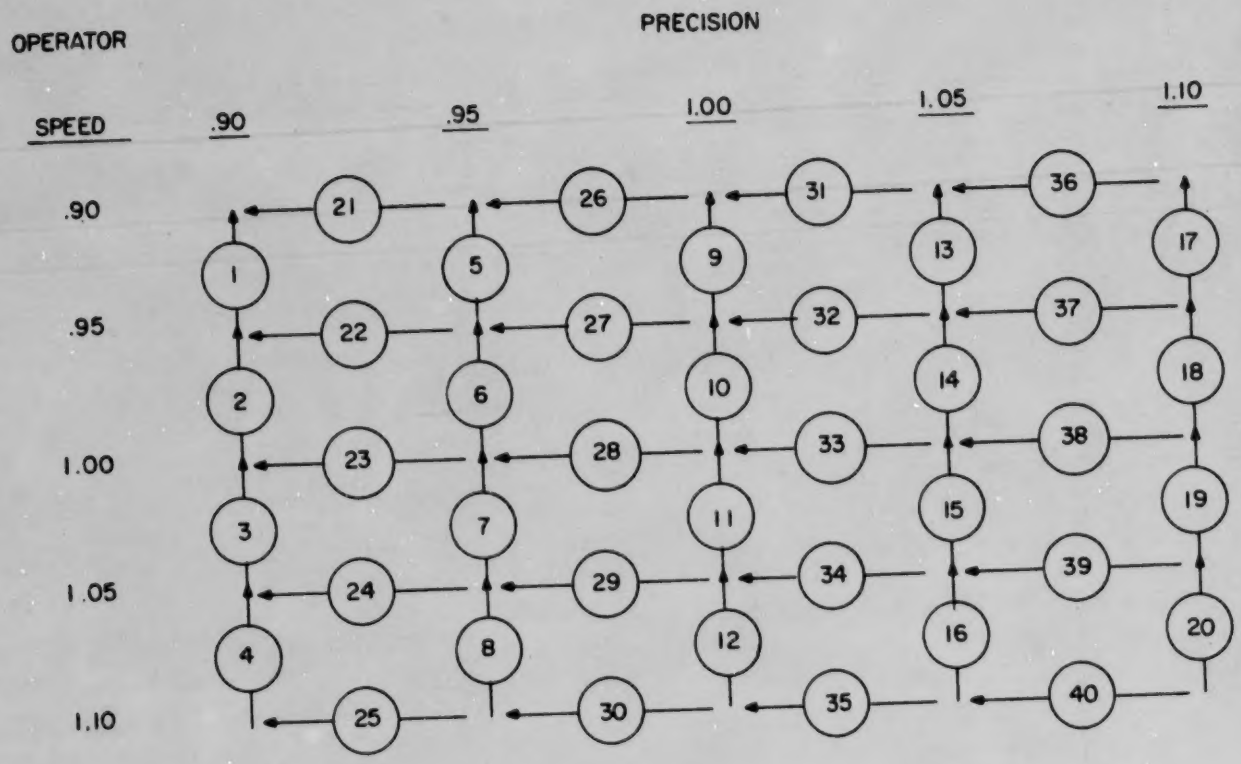


Figure 2-1. Operator level of proficiency predictions.

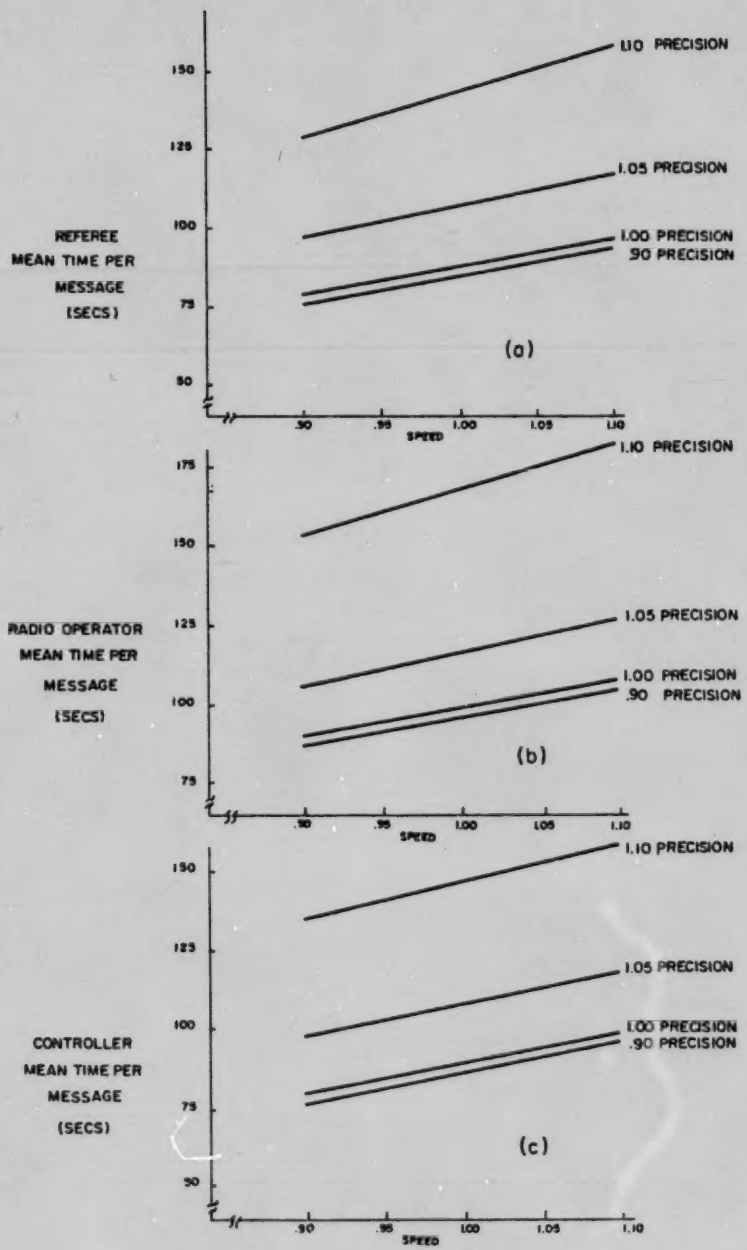


Figure 2-2. Mean time per message as a function of referee (a), radio operator (b), and controller (c) speed and precision.

some or all of a task. Improvement of precision below 1.0 had very small effects but quite significant changes in message processing time were produced when precision values were above 1.0. Mean time per message increased 60 to 70 percent as a result of a 10 percent decrement in precision above the nominal value (1.00), and about 10 percent for the corresponding decrease in operator speed below the nominal.

The effect of operator speed and precision on two of the four effectiveness components, responsiveness and thoroughness, was also marked. Figure 2-3 (a) and (b) displays this affect for the system thoroughness and responsiveness indices. The thoroughness measure is the ratio of the number of message blocks completed during an hour by a controller to the number of messages arriving during the hour. Responsiveness is determined as the average message processing time (referee start to controller end) divided by the average total elapsed message time.

Percentage Time Busy

The effect of operator speed and precision on percentage time busy for the referee and radio operator is presented in Table 2-3. Table 2-3 shows a strong trend for the percentage time busy to increase as the simulated personnel became slower and less precise. Regardless of speed or precision value, the simulated controller was busy 100 percent of the time. Percentage time busy for the referee varied between 14 and 30 percent over the range of values simulated; for the radio operator the range was 16 to 34 percent.

Level of Aspiration

The effect of level of aspiration of the simulated personnel depends, in the model, on their level of performance and the stress level. There are four conditions which affect performance:

1. the simulated personnel are performing below their level of aspiration and are only mildly stressed. The result is a positive pace adjustment factor due to aspiration.

2. the simulated personnel are performing above their level of aspiration and stress is low. This will tend to cause the operators to decrease their level of aspiration.

3. the simulated personnel are performing below their level of aspiration but are under high stress. In this case, the level of aspiration tends to be reduced and the speed of performance is degraded.

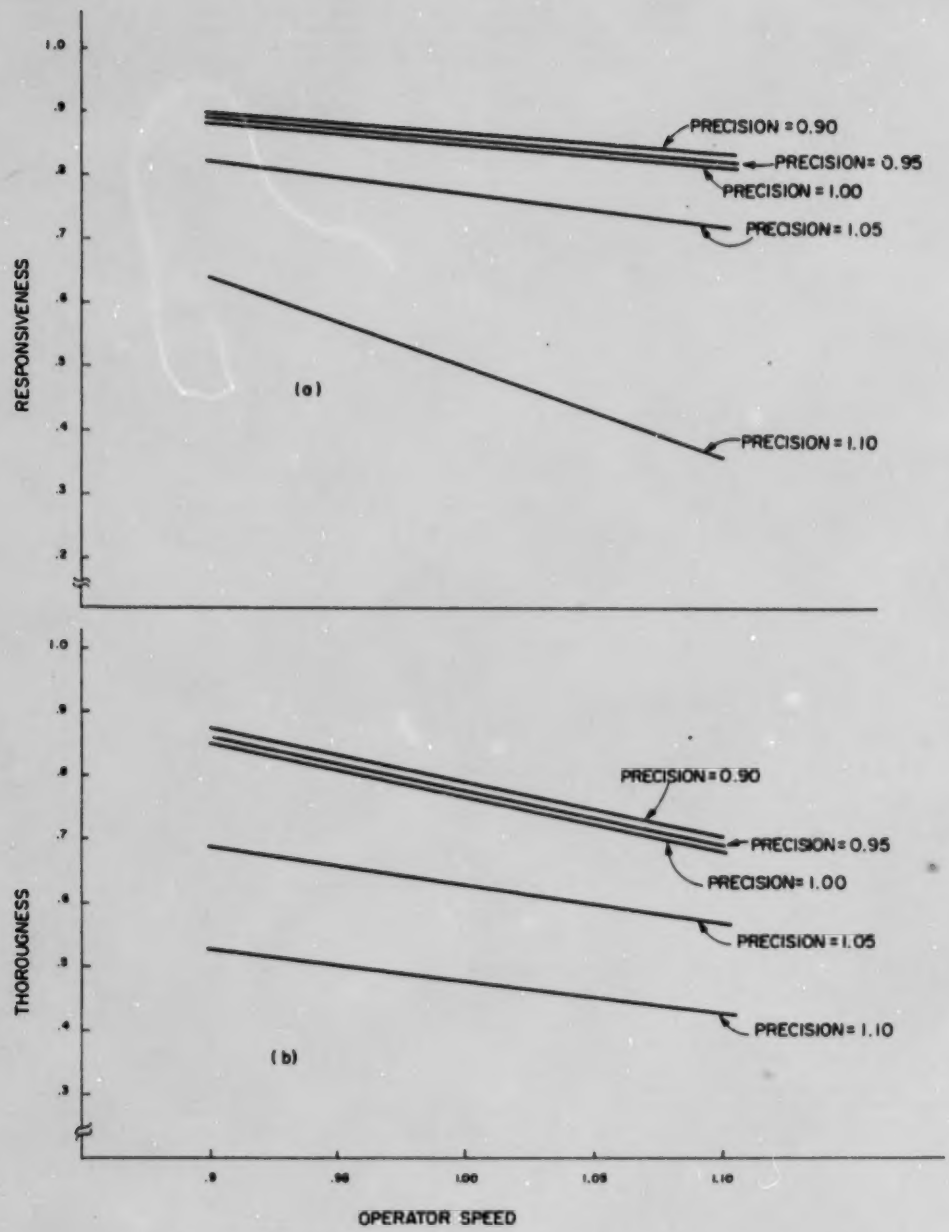


Figure 2-3. System responsiveness (a) and thoroughness (b) as a function of simulated personnel speed and precision.

Table 2-3

Percentage of Time Busy as a Function of Operator Speed and Precision

Speed		Precision .90	.95	1.00	1.05	1.10
.90	Referee	15	14	14	18	24
	Radio Op.	17	17	16	20	28
	Controller	100	100	100	100	100
.95	Referee	14	15	15	19	25
	Radio Op.	16	17	17	22	29
	Controller	100	100	100	100	100
1.00	Referee	16	16	16	20	27
	Radio Op.	18	18	18	23	31
	Controller	100	100	100	100	100
1.05	Referee	16	17	17	21	28
	Radio Op.	19	19	19	23	32
	Controller	100	100	100	100	100
1.10	Referee	17	17	18	22	30
	Radio Op.	19	20	20	24	34
	Controller	100	100	100	100	100

4. the personnel are performing at their level of aspiration but are under high stress. The result will be a slight reduction in stress.

The sensitivity test results indicated no noticeable affect by level of aspiration values over the .90 to .97 range on model outputs such as the effectiveness components or performance. The interrelations are such that the initial level of aspiration values are believed to be short lived in the model and dominated by other conditions. That is, regardless of the initial setting of aspiration values in the range, a final aspiration value of .96 was reached before the end of the first mission hour, after which initial aspiration values exerted no affect.

Stress Threshold

Within the NETMAN model, each simulated person's stress threshold is a variable which affects his response to stress. As stress increases, performance improves up to a point called the stress threshold. When stress passes the threshold, the facilitation is removed. Stress, within the model, is a function of message processing workload (i.e., percentage time worked). Stress begins to build beyond a nominal value of 1.0 when operators are working more than 66.6 percent but less than 95 percent of the time. Run 31 through 34 results displayed almost no affect of the operator stress threshold parameter alone over its tested range of values from 2 to 3. The results are shown in Table 2-4. Table 2-4 also indicates that the percentage of time the operators and computer were busy did not change as the stress threshold was increased from 2 to 3 on the stress tolerance scale. The threefold increase in percentage busy time due to a threefold increase in messages per hour input was insufficient to generate a stress condition because the percentage remained below 66 percent for the referee and radio operator and above 95 percent for the controller. The marked affects of changes in workloads, i.e., number of messages to be processed per unit time, are analyzed and presented in the section of this chapter which deals with message load effects.

Fatigue

Two factors determine fatigue as simulated in the NETMAN model. During each simulated work day, a performance decrement representing effects of

Table 2-4

Effects of Stress Threshold on Simulation Results

	Run Information			
Run Number	31	32	33	34
Stress Threshold	2	2	3	3
Messages per Hour	5	15	5	15
	Results			
Mean Time per Message				
Referee	85	86	85	85
Radio Operator	98	99	97	98
Controller	89	88	88	88
Effectiveness Component				
Thoroughness	0.75	0.25	0.75	0.25
Completeness	0.97	0.97	0.96	0.97
Responsiveness	0.85	0.85	0.86	0.86
Accuracy	0.99	0.99	0.99	0.98
Percent Time Busy				
Referee	16	48	16	48
Radio Operator	18	54	18	55
Computer	15	43	15	44
Controller	100	100	100	100
Final Stress, Each Operator	1	1	1	1

fatigue is applied with increasing influence over a 10 hour period. Second, on succeeding work days, the decrement is applied earlier in the work day. For the purpose of determining the sensitivity of the model's fatigue effects, three fatigue conditions were simulated. The three conditions were: first day of work, fifth day of continuous work, and ninth day of continuous work. Within each day, a 10 hour work period was simulated with workload in each hour held constant. Within each of three conditions, decrement was predicted to increase over the 10 hour working period. Moreover, the performance on the fifth day was anticipated poorer than on the first simulated day, and performance on the simulated ninth day was predicted to be worse than performance on the fifth day.

The principal results of the tests of model sensitivity to the fatigue variable are shown in Table 2-5. Operator performance, as measured by changes in the average time required per operator to process a message, degraded (i.e., time increased) as expected with longer mission periods. The percentage time increase from hour 1 to hour 9 were:

	MISSION DAYS		
Operator Type	1	5	9
Referee	9.3	14.0	19.8
Radio Operator	8.2	13.3	19.4
Controller	10.1	14.6	19.3
Average Percentage Time Increase	9.2	14.0	19.5

To evaluate further the impact of fatigue, two correlational analyses and an analysis of variance were completed for each operator type. First, product moment correlation coefficients were calculated for each of the three operator types and for three time periods (day 1, 5, and 9) by hour. The resultant values, shown below, indicate a high correlation between hour of work and simulated operator processing time per message. One would anticipate that as time on the job increased (fatigue increase), the processing time would also increase. Each correlation is based on an N of 10 hours.

Day	Referee	Radio Operator	Controller
1	.85	.84	.72
5	.94	.96	.89
9	.98	.97	.97

Table 2-5

Effects of Fatigue Variable on Mean Times per Message and on Thoroughness

	Day 1				Day 2				Day 3			
HOUR	Referee	Radio Op.	Controller	Thoroughness	Referee	Radio Op.	Controller	Thoroughness	Referee	Radio Op.	Controller	Thoroughness
1	86	98	89	.76	86	98	89	.76	86	98	88	.76
2	86	98	88	.77	85	98	87	.75	85	97	86	.78
3	85	98	88	.74	85	98	87	.75	87	100	91	.69
4	84	97	88	.75	85	99	89	.75	89	103	91	.72
5	86	97	86	.75	88	100	87	.79	92	102	91	.75
6	87	98	88	.77	90	103	92	.72	93	105	96	.67
7	87	100	89	.72	91	105	92	.71	95	108	97	.70
8	90	103	91	.76	92	108	96	.71	97	111	101	.68
9	92	103	93	.73	96	107	97	.69	100	110	101	.65
10	94	106	98	.67	98	111	102	.65	103	117	105	.64
Means	88	100	90	.74	90	103	92	.73	93	105	94	.71

An analyses of variance was completed for each operator type using the time per message per operator data of Table 2-5. Here the mission time was divided into three levels of days worked, and the 10 hourly results were divided into three periods of three hours each (the first hour was not used). The summary of the analysis is shown in Table 2-6. Statistical significance at the 0.001 level of significance was observed for all three operator types in both the effect of fatigue on performance for the mission time periods and for the hourly (day) periods. This indicates the clear and significant impact of fatigue on operator performance.

The effectiveness measure called thoroughness also showed a consistent reduction in operator performance over time. For the 10 hour simulated mission, the decrease in this measure was almost 16 percent.

Number of Networks

A network is composed of radio operator/referee teams who report to a common controller. The model can simulate as many as three such networks simultaneously. All networks are processed through the same simulated central computer and the resulting time sharing has an affect on message processing time. For the sensitivity test of the effect of number of networks, three levels were used--one, two, and three networks. A fairly heavy workload was imposed on the system in order to draw out the effect of the number of networks on overall processing time. It was predicted that increasing the number of networks would increase the time required to transmit to the computer and that increased waiting time within the computer would result from the increase in the number of networks being handled simultaneously. Within the simulation, each network had the same workload. Accordingly, the workload factor was held constant.

These predications were only partially fulfilled under the conditions of runs 14, 38, 39, and 40. Table 2-7 shows a negligible effect of number of networks on: the effectiveness components, mean message processing time, and proportion of time busy. Table 2-7 also shows the reason for this small effect. In the section of Table 2-7 dealing with proportion of time busy, the computer is predicted to be busy only 7 percent of the time for one network, 14 percent for two, and 22 percent for three networks. Accordingly, the workload assumed for the computer in these runs did not approach an overload situation. This effect was confirmed in field tests, described in Chapter III of this report, in which the computer was never a bottleneck in message processing. In terms of

Table 2-6

Summary of Analyses of Variance of Fatigue Effects

Operator	Source	Sum of Squares	Degress of Freedom	Mean Square	F	Level of Significance (p)
Referee	Total	700.7	26			
	Time (T)	464.3	2	232.15	64.59	< .001
	Day (D)	141.6	2	70.80	19.70	< .001
	T x D	30.1	4	7.60	2.11	
	Error	64.7	18	3.59		
Radio Operator	Total	732.96	26			
	Time (T)	444.74	2	222.37	40.84	< .001
	Day (D)	156.52	2	78.26	14.37	< .001
	T x D	33.70	4	8.42	1.55	
	Error	98.00	18	5.44		
Controller	Total	776.74	26			
	Time	474.30	2	237.15	35.97	< .001
	Day	140.74	2	70.37	10.67	< .001
	T x D	43.03	4	10.76	1.63	
	Error	118.67	18	6.59		

Table 2-7

Effects of Number of Networks Parameters on Simulation Results

	Run Information		
Run Number	38	39/14	40
No. of Networks	1	2	3
No. of Referees	8	16	24
No. of Radio Operators	8	16	24
No. of Controllers	1	2	3
No. of Computers	1	1	1
Effective Measures	**Results**		
Thoroughness	.79	.77	.76
Completeness	.97	.96	.97
Responsiveness	.86	.85	.86
Accuracy	.99	.99	.99
Overall	.90	.89	.89
Messages Completed Per Hour			
Referee	1081	1073	583
Radio Operator	2132	2113	1655
Controller	3243	3214	2487
Mean Time Per Operator Per Msg.			
Referee	86	86	84
Radio Operator	98	98	97
Controller	86	87	87
Proportion of Time Busy			
Referee	.16	.16	.16
Radio Operator	.18	.18	.18
Controller	1.00	1.00	1.00
Computer	.07	.14	.22

*Note: Messages completed is a total across four hours, five iterations, and eight referees per network.

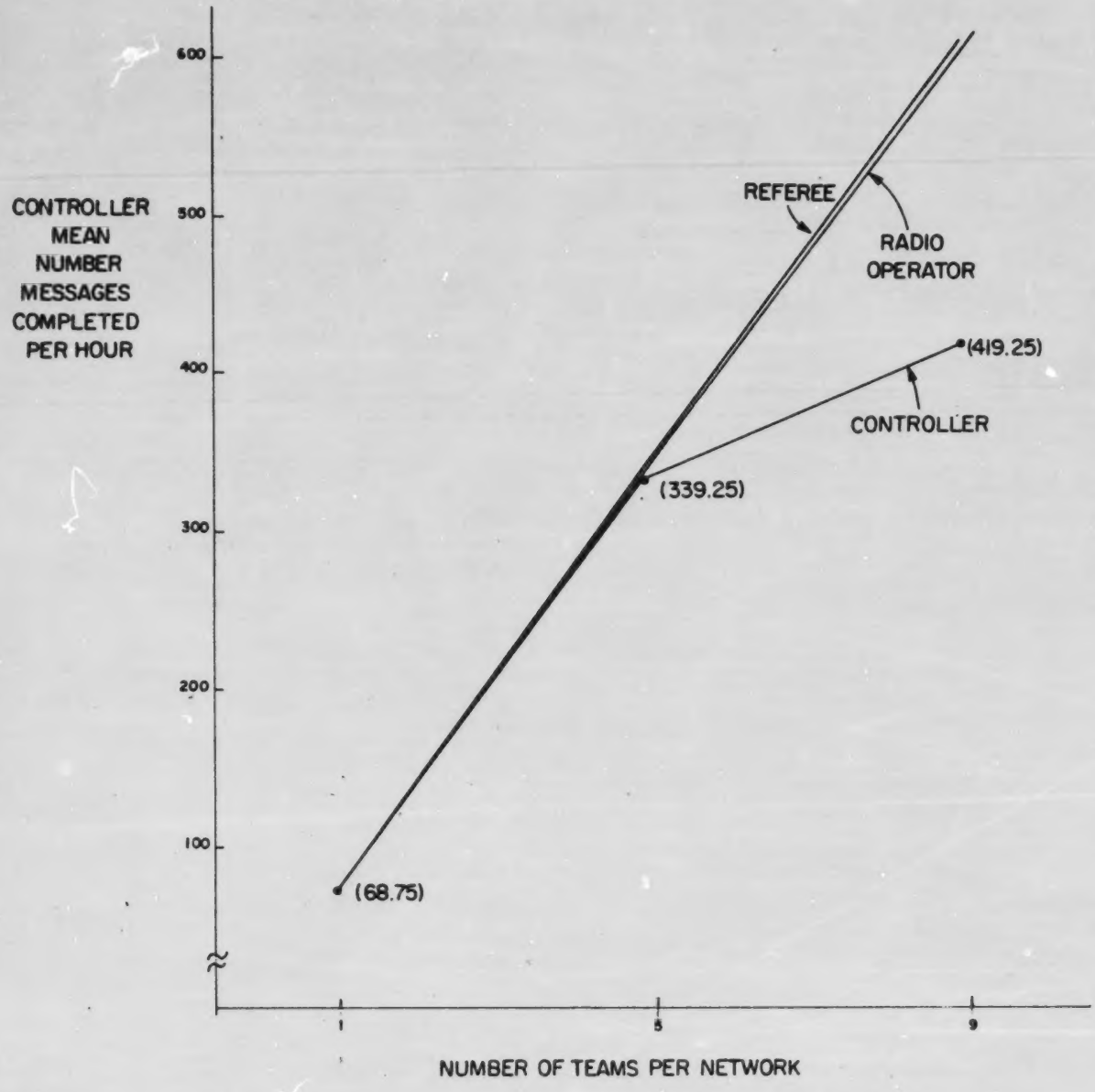


Figure 2-4. Messages completed as a function of number of teams per network.

Table 2-8

Summary of Analyses of Variance for Network Size Effect on Message Completion

Operator	Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Level of Significance
Controller	Total	272,110.25	11			
	Between	269,894.00	2	134,947.00	548.01	$p < .001$
	Within	2,216.25	9	246.25		
Referee	Total	581,918.67	11			
	Between	580,077.20	2	290,038.60	1,417.53	$p < .001$
	Within	1,841.47	9	204.60		
Radio Operator	Total	578,080.90	11			
	Between	575,752.67	2	287,876.34	1,112.80	$p < .001$
	Within	2,328.25	9	258.69		

Table 2-9

Summary of Analysis of Variance for Network Size Effect on Time

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Level of Significance
Total	16,485,944.79	20			
Between	16,395,512.00	2	8,197,755.90	1,631.70	$p<.001$
Within	90,432.25	18	5,024.00		

messages processed, however, adding networks produced an effect directly proportional to the number of networks added. That is, two networks processed about twice as many messages as one network, and three networks processed about three times as many messages as one network. This proportional relationship holds true at the referee, radio operator, and controller levels of the system.

Number of Teams per Network

Within the model, up to nine radio operator/referee teams may be assigned to each controller. An increased number of teams could be expected to increase the number of messages processed. Three levels of teams per network were tested: one team per network, five teams per network, and nine teams per network.

Figure 2-4 shows the effect of the three team levels on the number of messages per hour which were completed by each personnel type per hour over the four hour mission simulated. As expected, increasing number of teams per network resulted in an increased number of messages processed. However, for the controller, the increase was at a declining rate. For one team network, the result was 68.75 messages per hour completely processed (i.e., processed by the controller). Through the five team situation there was essentially no degradation since these teams processed 339.25 messages per hour--only about 1 percent below the single team rate. But, in the five to nine team range, the controller became a bottleneck. For the nine team situation, only 419.25 messages were processed as compared with the potential (9 x 68.75 = 618.75). The other operators' message processing was almost linear with the number of teams per network; a one-way analysis of variance (Table 2-8) for three networks was calculated independently for each operator type. For all three types, differences between number of networks showed significance at the .001 level.

Figure 2-5 displays the relationship between the number of teams per network and the logarithm of mean time involved in the reception of messages by the controller. The between teams variance (1, 5, 9 teams) was determined to be statistically significant ($p < .001$). Table 2-9 summarizes the pertinent analysis of variance. The data used in this analysis comprised the time spent in segment 15, which corresponds to delay to controller. Since there are seven message types and there was no difference between message types, there was an N of seven within each cell.

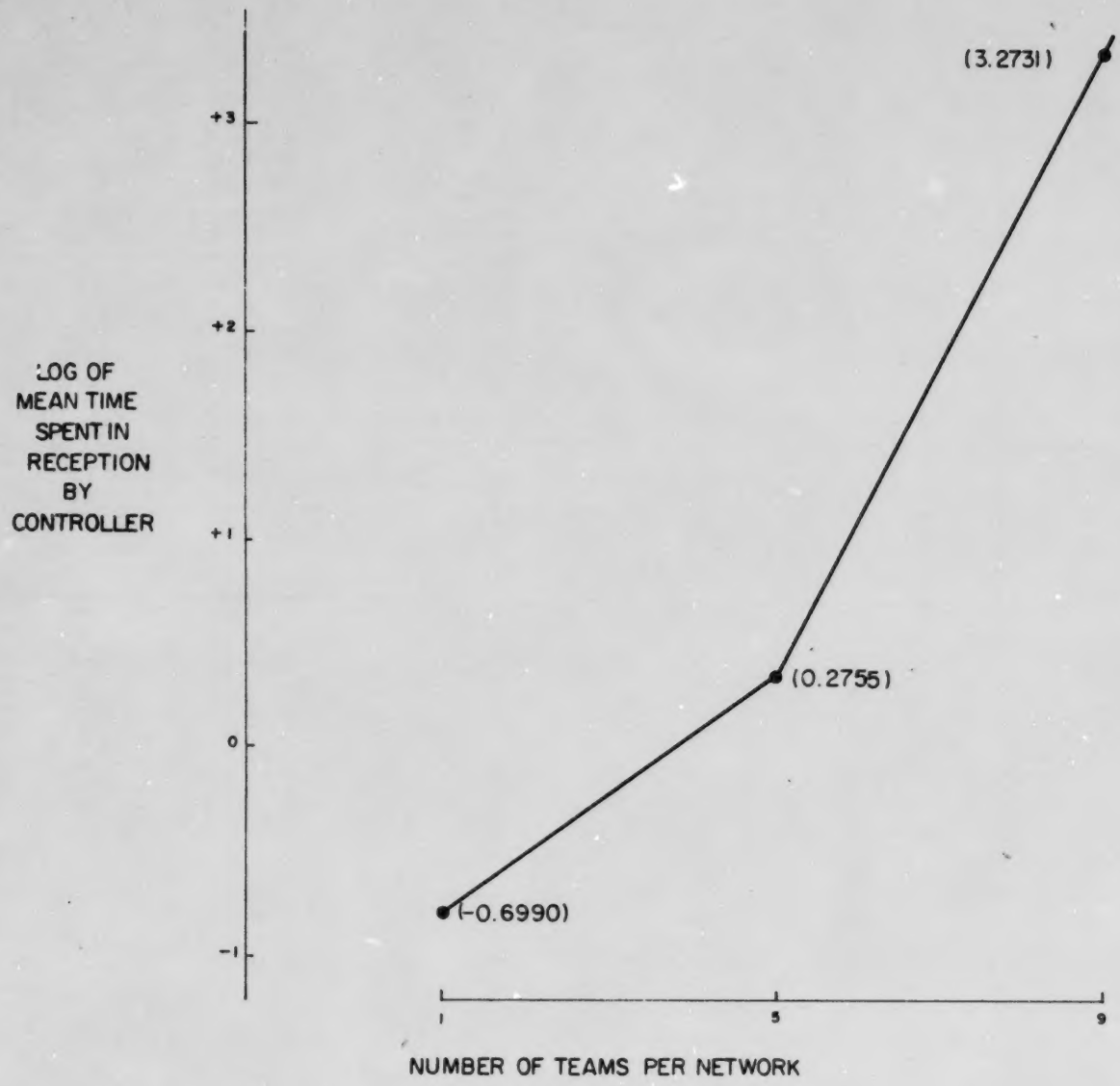


Figure 2-5. Effect of number of teams on time spent by controller on message processing.

Table 2-10

Effect of Number of Teams Per Network on Effectiveness Measures

Number of Teams Per Network	Effectiveness Measure				
	Thoroughness	Completeness	Responsiveness	Accuracy	Overall
1	.92	.97	.86	.99	.93
5	.96	.97	.86	.99	.94
9	.68	.97	.86	.99	.87

Table 2-11

Summary of Analysis of Variance for Undetected Error Probability Effect

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Level of Significance
Total	1.959	11			
Between	1.958	2	.9790	9790.	p <.001
Within	0.001	9	.0001		

For the one team per network case, 0.2 seconds were spent in message reception out of a total message time of 285 seconds. This increased to an average of 0.8 seconds for the five team per network case out of a total time of 288 seconds. However, in simulation run 43 (nine teams per network), the average time spent by the controller in message reception was 1875 seconds out of a total of 2172 seconds. Again, this suggest overloading of the controllers for this case.

Table 2-10 shows that only one of the effectiveness measures, thoroughness, was impacted by variation of the number of teams per network. After the overload condition set in, the anticipated reduction in thoroughness was indicated.

Undetected Error Probability

Many of the errors which occur in the simulated message handling are detected and corrected; others pass through and contaminate the simulated data base. The probability of an undetected error is specified by NETMAN model input in terms of the probability of a low importance error and in terms of the probability that a significant error will enter the data base. These error probabilities are expressed in terms of probability per message. These probabilities are used in the model in the calculations of information loss and effectiveness. The error probabilities selected for test were: .01, .1, and 1.0 (i.e., one message in 100 will contain undetected errors, one message in 10 will contain undetected errors, and one undetected error for every message processed.)

Figure 2-6 shows a plot of the resultant effectiveness measure Accuracy = $(1 - \frac{\text{total information loss}}{\text{messages completed}})$ as a function of undetected error probability. The other three effectiveness components were essentially unaffected by the undetected error probability over the range .01 to 1.0; the resulting output ranges were:

Thoroughness	=	.74 - .77
Completeness	=	.96 - .97
Responsiveness	=	.85

A one-way analysis of variance indicated the effect on accuracy to be statistically significant at the 0.001 level. The analysis is summarized in Table 2-11.

Message Frequency

The frequency with which the simulated referees send messages, whether originated by themselves or by the controller, is a prime

Figure 2-6. Effect of undetected error probability on the accuracy component of effectiveness.

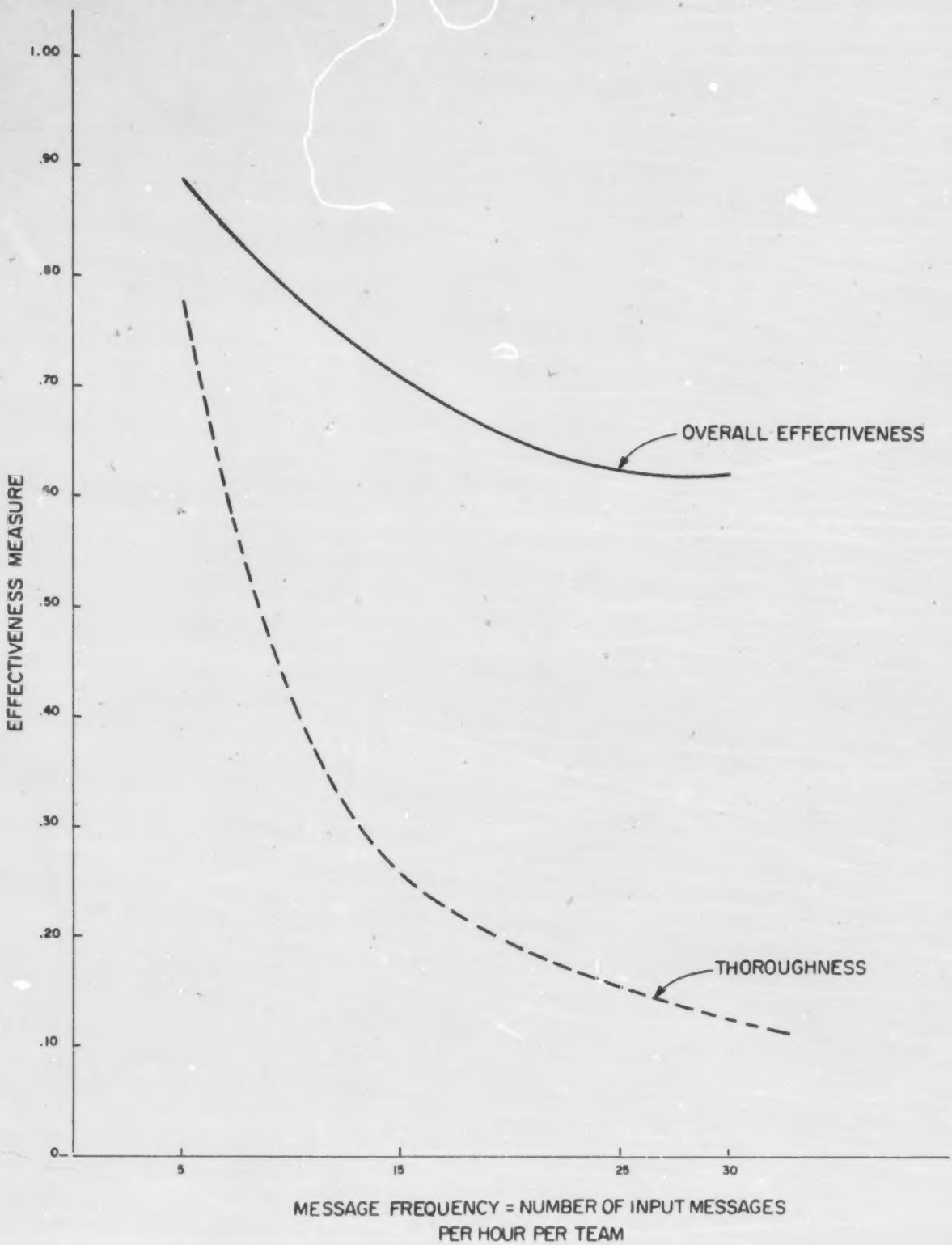


Figure 2-7. Effect of number of messages per hour on overall effectiveness and on effectiveness components.

Table 2-12

Summary of Analysis of Variance of Message Frequency Effect on Overall Effectiveness

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Level of Significance
Total	0.1806	15			
Between	0.1803	3	0.0601	2,404	$p < .001$
Within	0.0003	12	0.000025		

Table 2-13

Summary of Analysis of Variance for Task Difficulty Effect on Messages Processed Per Hour

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Level of Significance
Total	29,734	11			
Between	29,295.5	2	14,647.75	300.64	$p < .001$
Within	438.5	9	48.72		

determinant of the workload on the entire system. Message frequency varied over four levels: 5, 15, 25, and 30 messages per hour per team. This range represents a variation from light workload, through heavy workload, to a work overload.

Figure 2-7 shows (solid line) the effect of input message load per unit time on the overall operator effectiveness. This relationship was shown below to be significant at the .001 level as a result of a two way analysis of variance (effectiveness measure by message frequency). This variance analysis is summarized in Table 2-12. This finding is largely due to the very high changes in the thoroughness component. Much smaller changes were noted in the other effectiveness components.

The delay time between computer processing and controller start as a function of message frequency is shown below:

Message Frequency	Controller Delay		Proportion of Time Busy		
	Mean	Sigma	Referee	Radio Operator	Controller
5	949	98	.16	.18	.14
15	11,333	225	.49	.56	.44
25	11,181	364	.69	.79	.71
30	11,663	574	.72	.82	.86

As anticipated, delay increased as workload increased and the proportion of time busy rose even more rapidly as a function of workload.

Field Message Length

The length of each message is specified in the model in terms of the number of characters it contains. The number of characters to be processed is directly related to the time required to process the entire message, assuming that the time to process each character is held constant. The levels of message length tested were: 10, 22, and 100 characters.

The expected increase was observed in message processing as a function of number of characters. Figure 2-8 presents the obtained results. The Pearson product moment coefficient of correlation between processing time (mean time per message) and message length was calculated to be .9996 for the radio operator and .9995

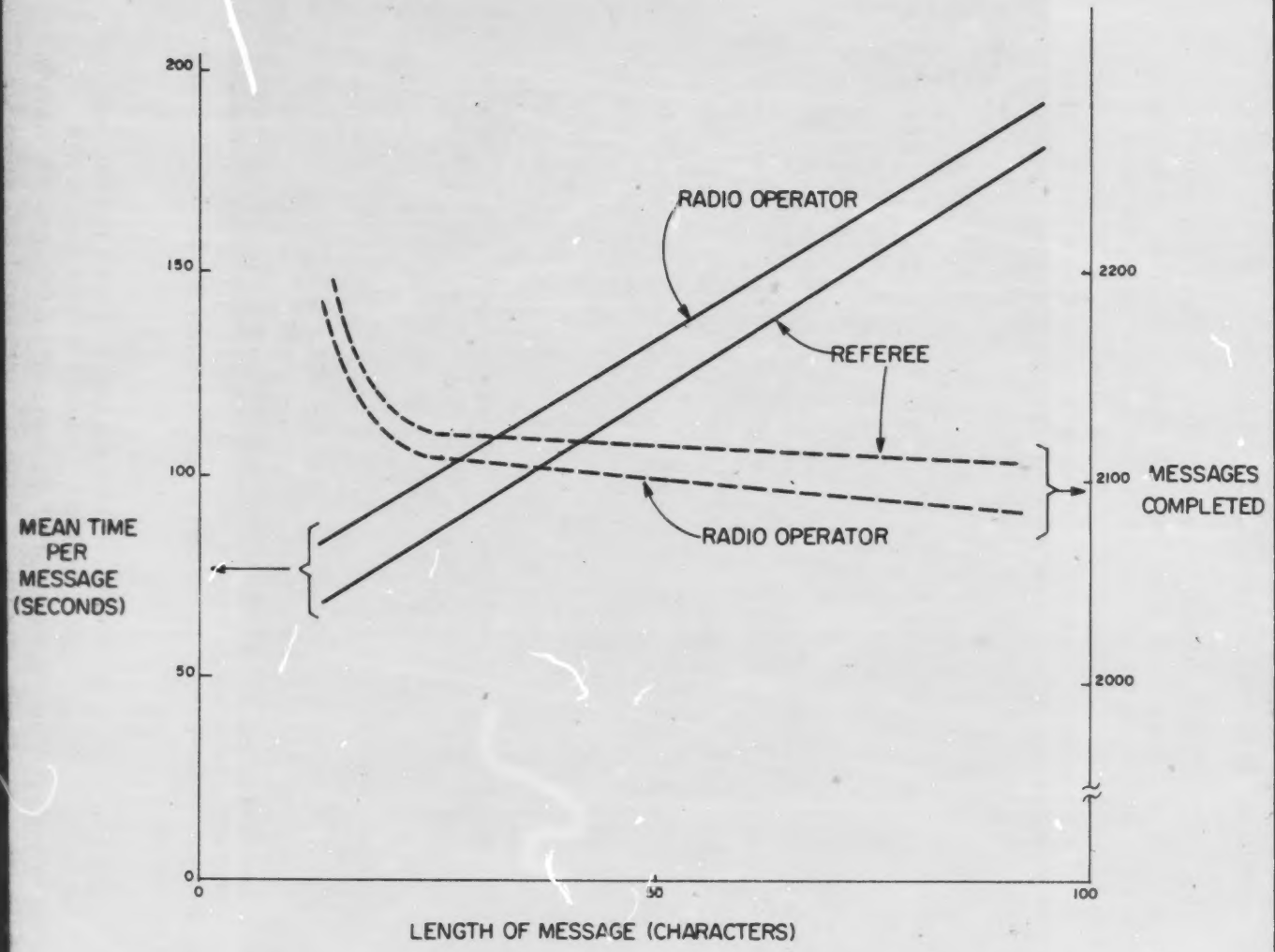


Figure 2-8. Effect of message length on messages completed and mean time per message for referee and radio operator.

for the referee. It was also confirmed that field message length had a marked effect on effectiveness components, particularly responsiveness and accuracy, as shown below:

Message Length	Thoroughness	Completeness	Responsiveness	Accuracy	Overall
10	.76	.96	.89	1.00	.90
22	.77	.96	.85	.99	.89
100	.76	.97	.47	.74	.72

Similarly, Figure 2-8 shows a corresponding affect, in the anticipated direction, of message length on messages completed. As message length increased from 10 to 100 characters, the number of messages completed decreased by about five percent.

Controller Message Length

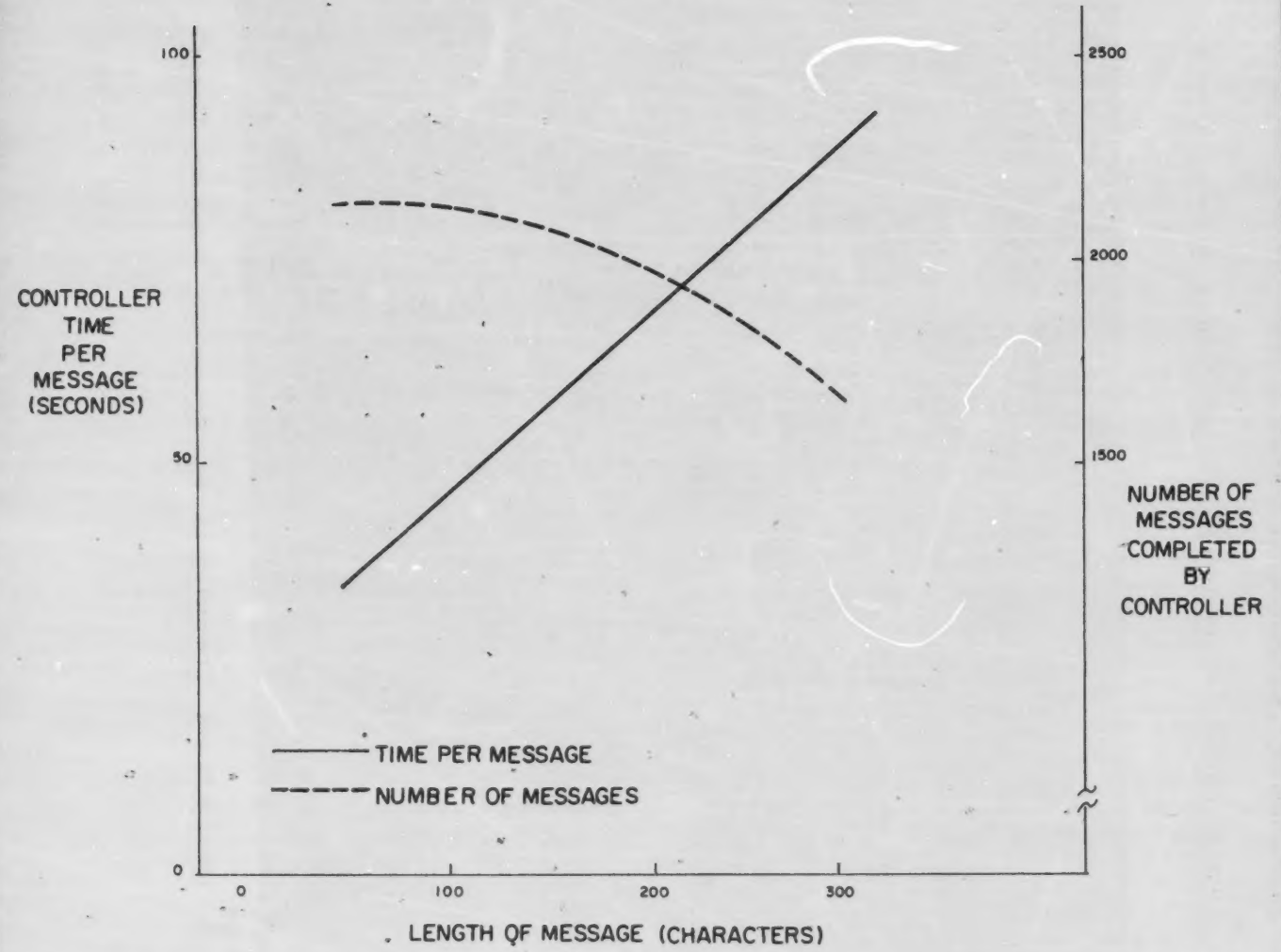
A message which is sent by the simulated field teams is automatically decoded by the simulated computer before it is displayed to the controller. Messages to the controller are typically longer, in terms of number of characters, than the messages composed in the field. For the purpose of these sensitivity tests, the range of message lengths used was 10, 22, and 100 characters.

The increase in mean time per message as a function of increased message length was as pronounced for the controller as it was shown to be for the other personnel. The pertinent data are displayed in Figure 2-9, and the corresponding triserial correlation between controller message length and mean time per message was .99.

Correspondingly, the total number of messages completed by the controller was reduced as message length increased, as shown in the dashed curve of Figure 2-9.

The effects of controller message length on the effectiveness components were:

Message Length	Thoroughness	Completeness	Responsiveness	Accuracy	Overall
50	.97	.97	.91	.99	.96
100	.98	.97	.90	.99	.96
300	.77	.96	.85	.99	.89



60

Figure 2-9. Effect of message length on messages completed and mean time per message for controller.

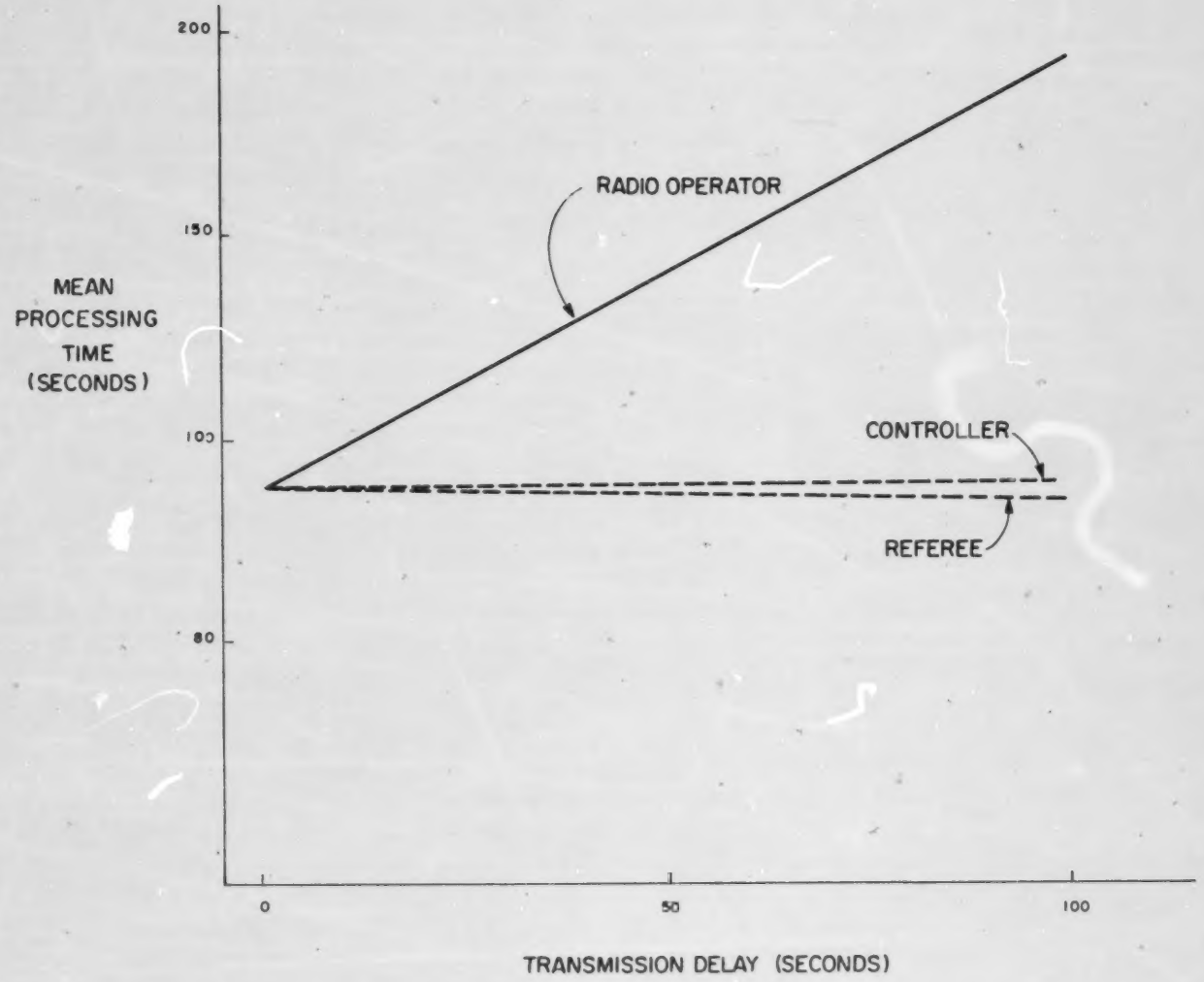


Figure 2-10. Effect of transmission delay on processing time.

Transmission Delay

Before a radio operator sends a message, he selects a transmission frequency and listens to determine whether or not the frequency is clear. Depending on the traffic, the message transmittal may be delayed. As predicted, simulating this delay added to the average message processing time for the simulated radio operator and caused a reduction in the number of messages which he handled per hour.

Figure 2-10 shows these effects for the three levels of delay in the sensitivity tests: no delay, 10 second delay, and 100 second delay. No effect was to be expected for the controller or the referee because this variable should only operate on the radio operator processing time. This consistency is also shown in Figure 2-10. The triserial correlation between radio operator processing time and the duration of the transmission delay was calculated to be 0.974. The effect of this delay on the effectiveness components is shown by the following data to be very modest -- less than a 5 percent change over the entire range of transmission delays.

Transmission Delay (secs)	Thoroughness	Completeness	Responsiveness	Accuracy	Overall
0	.76	.96	.87	.99	.89
10	.77	.96	.85	.99	.88
100	.76	.97	.72	.99	.85

Task Difficulty

Task (mission) difficulty within the model is indicated by the value of task element success probability assigned to each of the tasks in the sequence for each person simulated. Typically, system personnel will correct the error which caused failure on a task element immediately. The additional time spent in correction and touch up, however, increases the time required for task completion and would be expected to reduce the overall performance effectiveness. Task element success probabilities were varied over three levels. Under the low difficulty condition, success probabilities were 0.99 on each task element. In the moderate difficulty condition, the success probability was 0.80, while in the high difficulty condition a success probability of 0.60 was entered for each task element. It was predicted that increased task difficulty would produce a strong decrement in the number of messages processed, an increase in the time per message, and a decrease in overall effectiveness.

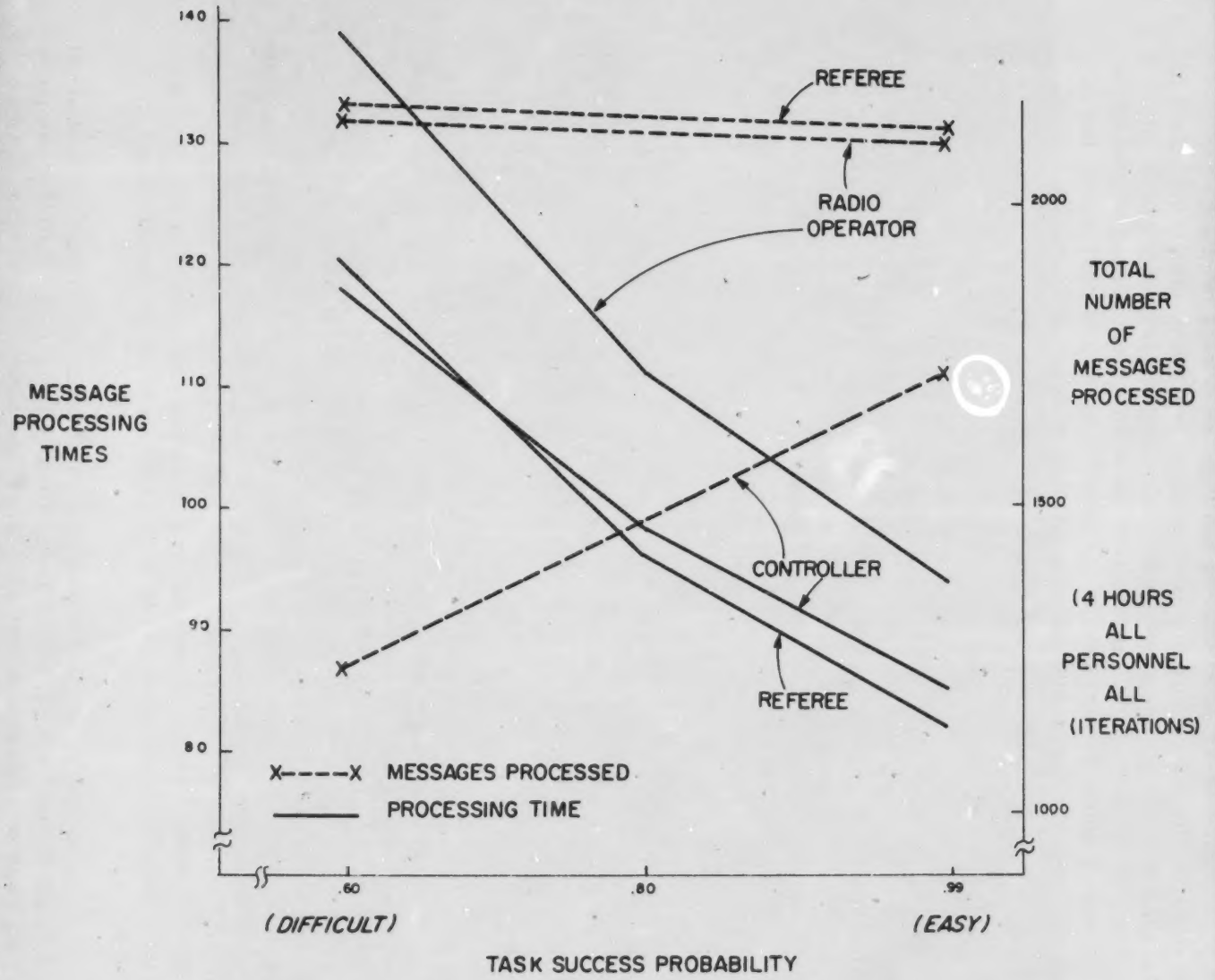


Figure 2-11. Effect of task element success probability on message processing times.

Figure 2-11 confirms the anticipated effects of changes in task difficulty on the number of messages per hour. Here the effect of increased messages processed as tasks become less difficult is clearly noted for all operators. In addition the number of messages processed increased about 29 percent for the controller who is overloaded but not at all for the referee and radio operators who are only mildly loaded.

A one-way analysis of variance (Table 2-13) indicated significance at the .001 level for messages processed per hour for the controller.

Figure 2-11 also shows the expected effect of decreased message processing times for the controller as the difficulty of the task increased. An analysis of variance (Table 2-14) also indicated this effect to be statistically significant. However, the referee and the radio operator showed virtually no change because these personnel were very lightly loaded. Also note that the total number of messages processed by the radio operators never quite equalled the number processed by the referees due to the sequence of message handling.

Several effectiveness components indicated degradation as difficulty increased (Figure 2-12). An analysis of variance calculated for the overall effectiveness measure indicated statistical significance ($p < .001$). This analysis is summarized in Table 2-15.

Cost to Run

The second consideration in the sensitivity tests was cost to run. Cost to run was examined from both the direct cost and computer time used point of view. The U1108 facility at Edgewood, like many computer facilities, charges different rates depending on the time of day of use. Computer runs made between 0800 and 1600 hours are most costly; use of the facility between 1600 and 2400 is discounted and use between 0000 to 0800 is further discounted. The rate is also a function of memory, input, and output requirements. Similarly "batch" runs are much cheaper than "demand" runs. A typical five iteration simulation in the "batch" mode cost $.45 and used 6.848 seconds of central processing unit (CPU) time.

A terminal set up for another run which also involved an aborted simulation run cost $2.88. The terminal was connected to the computer from 1046 to 1132 hours and used 37.47 seconds of CPU time. As an example of one of the higher costs of running the model, a simulation was run and both lengthy set up and print out were required. This run was made from a terminal which was on line with the computer from 1411 to 1556 hours. This run cost $24.79 and required 65.63 seconds of CPU time.

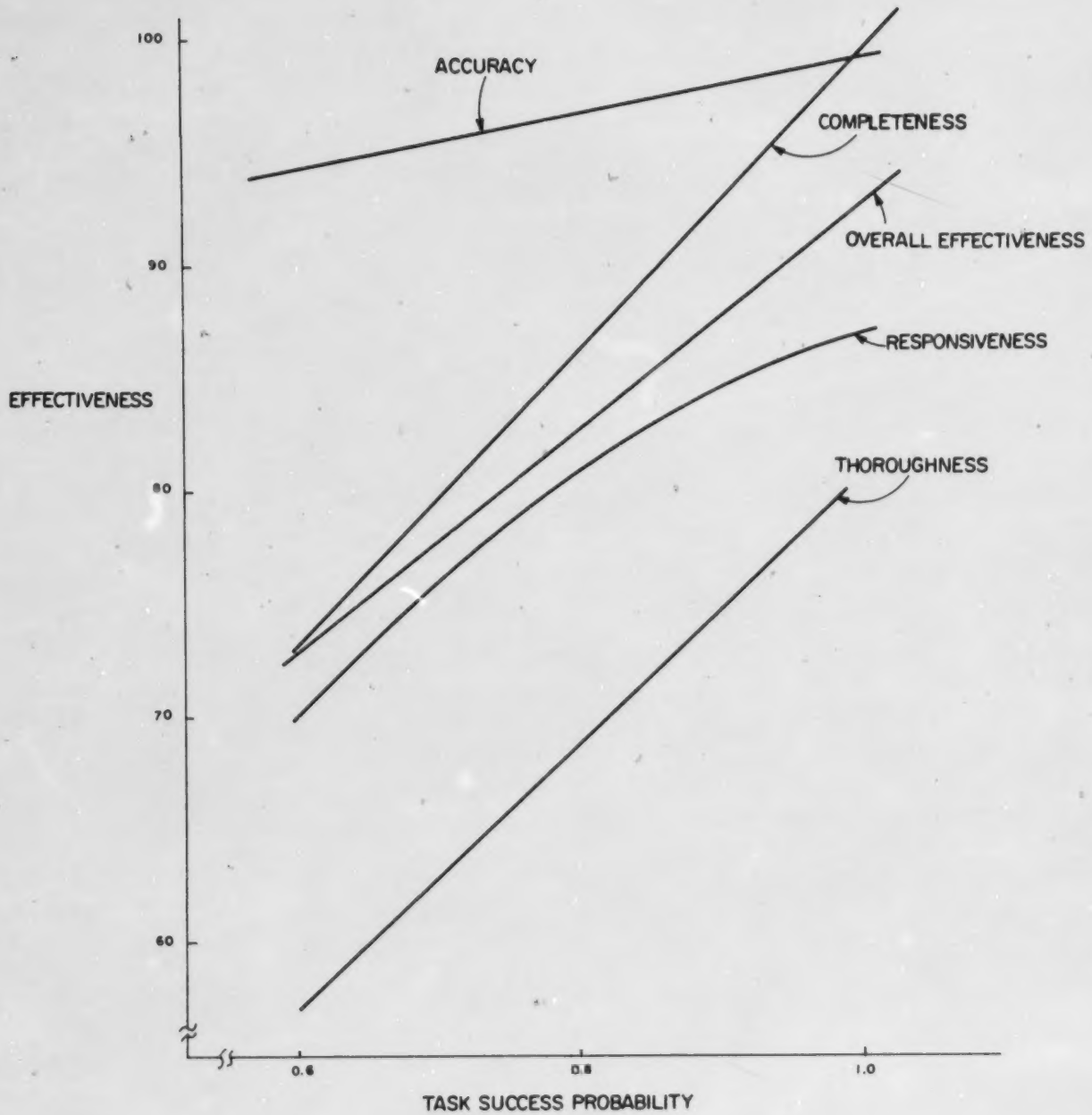


Figure 2-12. Effect of task element success probability on effectiveness elements and on overall effectiveness.

Table 2-14

Summary of Analysis of Variance for Task Difficulty Effect on Message Processing Time

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Level of Significance
Total	2,384.67	11			
Between	2,344.67	2	1,172.33	264.04	$p < .001$
Within	40.00	9	4.44		

Table 2-15

Summary of Analysis of Variance for Task Difficulty Effect on Overall Effectiveness

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Level of Significance
Total	0.691	11			
Between	0.688	2	.0344	1,033.03	$p < .001$
Within	0.0003	9	.0000333		

Small program or file changes typically required only a few seconds of CPU time and cost less than $.25.

Due to the size of the program, long delays were often experienced while waiting for sufficient computer space to become available. In the case of batch runs, this delay was as high as several hours. In the case of terminal runs, when delays exceeded a few minutes, the run was aborted due to the relative preponderance of telephone line costs compared to computer costs.

Ease of Use

Although the NETMAN model cannot be called "easy to use", the input is organized in a systematic manner which promotes ease of use. Some complexities in the preparation of the input data are described in this section.

An example of the volume of data required is shown in data prepared for the D-Day simulation run described in the next chapter of this report. This input required 92 cards including over 600 values.

The NETMAN model requires fixed format entry. That is, where data are called for, the exact card column(s) or spacing on a terminal is necessary. Any deviation from the indicated spacing will cause either a discrepancy in the input or, more likely, the program will reject the input as invalid.

The computer program does not double check the consistency of the input data. For example, if two networks are specified in the input, the computer will read in the number of men in each from the appropriate card columns. If entries in these columns are omitted, a zero is entered. This will cause the program to terminate execution once the simulation has begun and it may be quite difficult to trace back to the source of the problem.

Although the organization of the initial input data to the NETMAN model can be quite time consuming, changing small amounts of input is quite easily accomplished in the experimenter interactive mode. Such variables as personnel characteristics and message characteristics can be temporarily changed and the effect on output determined. Normally, a great many variations on the input data are run in order to

study the range of the simulated system's performance. Accordingly, the time for initial input data preparation, when amortized over a number of runs under various conditions, is not excessive.

Program Error

Two "errors" were found during the sensitivity tests. These "errors" provide hitherto unknown limitations of the NETMAN program. First, the model cannot be used to simulate a full 12 hour mission as was initially intended. A 10 hour mission was simulated without problem. Secondly, a message load over 30 messages per hour per referee cannot be handled. It is believed that both of these limits are due to the present size of data storage arrays. In extreme cases the program execution is automatically terminated by the computer, probably due to attempts to make illegal transfers of data. The changes required to remove these limitations are believed to be relatively minor.

III. MODEL VALIDATION

Field validation is another step in completing the NETMAN development cycle. The purpose of the validation is to derive indices of the model's precision in estimating the system and human effects it was designed to predict. A field exercise management environment from which appropriate criterion data might be derived, i.e., one employing information processing characteristics similar to those simulated by the netman model, is essential to such validation. The Marine Corps' Tactical Warfare Simulation, Evaluation, and Analysis System was considered to be an ideal system for validation purposes. TWSEAS' predecessor, TWAES, was one of the systems from which the NETMAN simulation was developed.

This chapter summarizes the planning for and collection of actual field data suitable for evaluating the validity of the NETMAN computer model. By way of overview, the validation was based on comparing simulation model results with direct observational data obtained from the Tactical Warfare Simulation, Evaluation, and Analysis System (TWSEAS). Data were collected at Camp LeJeune, North Carolina.

Criterion Data Attributes

The purpose of the present validation was to compare the model's predictions with the performance of an actual system. When such a validation is performed, the actual performance data, called criterion data, must be carefully collected and must represent as close an approximation as possible to the situation that model was built to simulate. The TWSEAS meets this requirement.

In evaluating the TWSEAS as a vehicle against which the NETMAN may be considered, the attributes of an "ideal" criterion must be held in mind. A list of such attributes is presented in Table 3-1.

Criterion Variables Selected

The eight TWSEAS criterion variables selected for assessing the validity of the NETMAN model are shown in Table 3-2, together with the associated model variables.

Table 3-3 presents a formulation whereby the relative goodness of the eight criteria may be determined. It shows a matrix of evaluation scores for each criterion against each of the criterion aspects which was identified in Table 3-1. The maximum total score that any one of the field criteria could achieve is 48 (columns of Table 3-3). No criterion received this score. The minimum possible score is 12. By this yardstick, no criterion received a minimal evaluation. Transmission delay

Table 3-1

Desirable Criterion Attributes

Criterion Attribute	Notes on Acceptability of Criteria
Data Availability	An acceptable criterion is one which is backed by substantial empirical data for the range of populations to be modeled.
Data Reliability	An acceptable criterion is one for which the error range of the available data is known and minimum.
Relevance to Situation	An acceptable criterion is one which is critical to and possesses obvious saliency to the acts and behaviors of individuals and groups involved in the simulation.
Sensitivity	An acceptable criterion is one which will vary as the result of various conditions expected to be changed during simulation runs.
Objectivity	An acceptable criterion is one which is based on objective information.
Uniqueness	A preferable criterion is one which is associated with unique output variance. Each variable should contribute to the richness and completeness of validation.
Suitability of Form	That criterion is preferred for which the available data do not require excessive transformation, rescaling, preprocessing, or translation.
Generality	That criterion is preferred which is applicable to a wide range rather than only a limited number of situations.
Comprehensibility	A criterion should be easily understood by the users of the model's output.
Utility	That criterion is best which is most important to the questions the planner and model's user wishes to ask.
Freedom from Triviality	A criterion which will reflect only minor effects should be avoided.
Heuristic Value	The comparison of criterion data with model output should raise questions relative to the model as well as to answer questions.

Table 3-2

Criterion Variable Description and Method of Measurement

Model Variable	Basis in Model	Basis in TWSEAS
Thoroughness	$\frac{\text{Number of messages processed}}{\text{Number of messages available for processing}}$	Count for each hour during exercise
Responsiveness	$\frac{\text{Message processing time (man hours)}}{\text{Total time (includes above plus all delays)}}$	Time measure during exercise
Overall effectiveness	Weighted equation combining thoroughness, responsiveness, accuracy, and completeness	Controller ratings after exercise
Total message processing time	Total message time from referee start until controller completion	Time measure during exercise
Referee message processing time	Mean referee time per message	Time measure during exercise
Radio operator message processing time	Mean radio operator time per message	Time measure during exercise
Controller message processing time	Mean controller time per message	Time measure during exercise
Transmission delay	Mean transmission delay time	Time measure during exercise

Table 3-3

Criterion Evaluation

Criterion Attribute	Total Score	Thoroughness	Responsiveness	Overall Effectiveness	Message Processing Time				
					Total	Referee	Radio Operator	Controller	Transmission Delay
Data Availability	29	E	G	F	E	E	E	E	E
Data Reliability	23	G	G	F	G	G	G	G	G
Relevance to Situation	28	E	E	E	E	G	G	G	G
Sensitivity	31	E	E	E	E	E	E	E	G
Objectivity	29	E	E	P	E	E	E	E	E
Uniqueness	29	E	E	G	G	E	E	E	G
Stability of Form	28	E	E	P	E	E	E	E	G
Generality	23	E	E	E	G	F	F	F	F
Comprehensibility	31	E	E	E	E	E	E	E	G
Utility	27	G	G	E	E	G	G	G	E
Freedom from Triviality	31	E	E	E	E	E	E	E	G
Heuristic Value	32	E	E	E	E	E	E	E	E
Totals		46	45	37	45	43	43	43	39

Score:

4 = E = Excellent
3 = G = Good
2 = F = Fair
1 = P = Poor

received a low evaluation score largely because it depends on equipment characteristics and transient effects. Overall effectiveness received the lowest rating due to its lack of objectivity and stability. Thoroughness and responsiveness received the highest scores.

From the point of view of attributes across criteria, heuristic value, comprehensibility, and freedom from triviality received the highest scores (32, 31, and 31 respectively), while data reliability and generality received the lowest scores (23 in each case). The possible score range for attributes was 8 to 32. Most attributes scored relatively high in this range.

Model Validation Concepts

The need to validate a computer simulation model was recognized early. Siegel and Wolf (1959) spoke about the need to investigate the properties of new models, to apply a new model to more than one test, to evaluate the effectiveness and consistency of a model, and about the desirability and feasibility of validating a model against outside criterion data. The desirability of model validation by comparison of model results against those obtained through a controlled laboratory test was also recognized early (Siegel, Wolf, & Sorenson, 1962). The suitability of the Campbell and Fiske (1959) technique for test validation of models was shown in Siegel, Lautman, and Wolf (1972).

Emshoff and Sisson (1970) in a discussion on model validity concluded: "the only possible evidence of validity for a simulation model that has been developed specifically for a situation is that the model has made satisfactory predictions in the past." They suggested five "preliminary criteria for evaluating first time models" as described by Hermann (1967). These five are identified by an asterisk in the more comprehensive list of 15 criteria for evaluating a simulation model which are displayed in Table 3-4. These criteria are not necessarily mutually exclusive. Some are overlapping, but all are considered important in some sense and/or for some classes of models.

In order to place these criteria into some perspective and to view the sequential steps through which a model might pass, consider Figure 3-1, which attempts to tie together the various model development/validation phases with these 15 criteria for model evaluation (per Table 3-4). Figure 3-1 displays the major steps (large rectangles) from concept and model requirements derivation through the situation in which a model can be considered for decision aiding and eventually for decision making. The 15 numbered vertical arrows, representing the 15 criteria, show that each step in the process yields some measure of utility, feasibility, cost, reasonableness or validity.

Table 3-4

Criteria for Evaluating the Utility of a Computer Model

Criterion	Definition
1. Internal consistency	Extent to which the constructs of the model are marked by coherence and similarity of treatment
2. Indifference to trivial aggregation	Potential of the model to avoid major changes in output when input groupings or conditions undergo insignificant fluctuations
3. Correct prediction in the extreme (predictive or empirical validity)	Extent of agreement (correctness of predictions) between model and actual performance at very high/low values of conditions
4. Correct prediction in mid range (predictive or empirical validity)	Like above for middle ranges values of conditions
5. Construct validity	Theoretic adequacy of the model constructs
6. Content (variable parameter) validity (Fidelity) *	Extent to which the model's variables/ parameters match real life conditions
7. Realism or "face validity" *	Extent to which selected content matches each attribute modeled
8. Richness of output	Number and type of output variables and forms of presentation
9. Ease of use	Extent to which an analyst can readily prepare data for, apply, and extract understandable results from the model
10. Cost of development	Value of effort to conceive, develop, test, document and support
11. Transportability-generality	Extent to which different systems, missions, and configurations can be simulated

Table 3-4 (continued)

12.	Cost of use	Value of all effort involving use of model including data gathering, input, data processing, and analysis of results
13.	Internal validity *	Extent to which outputs are repeatable when inputs are unchanged
14.	Event or time series validity *	Extent to which simulation predicts events and event patterns
15.	Hypothesis validity *	Extent to which model relationships correspond to similar relationships in the observable universe

* Approaches to validation defined by Hermann (1967)

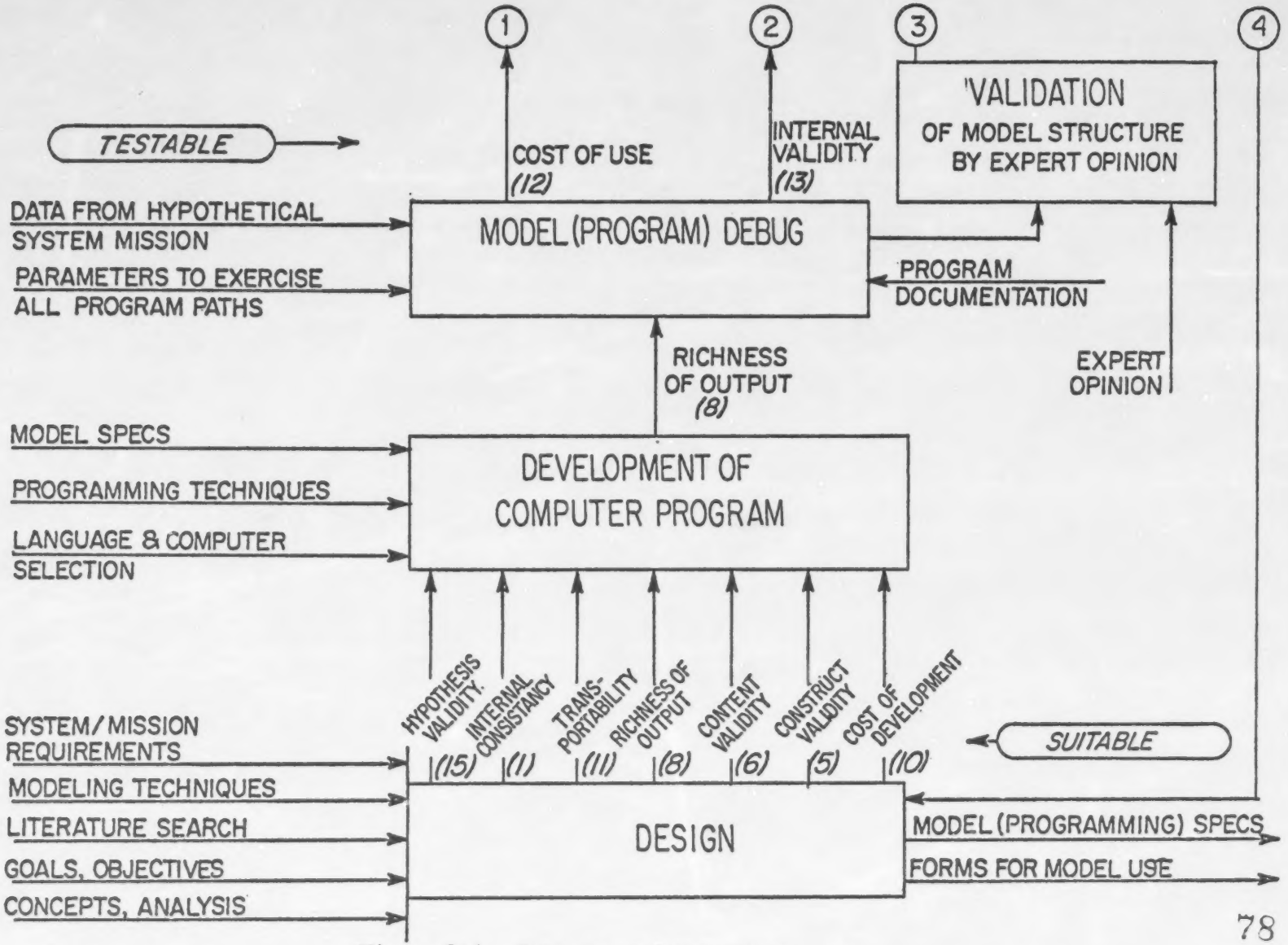


Figure 3-1. Steps in model development.

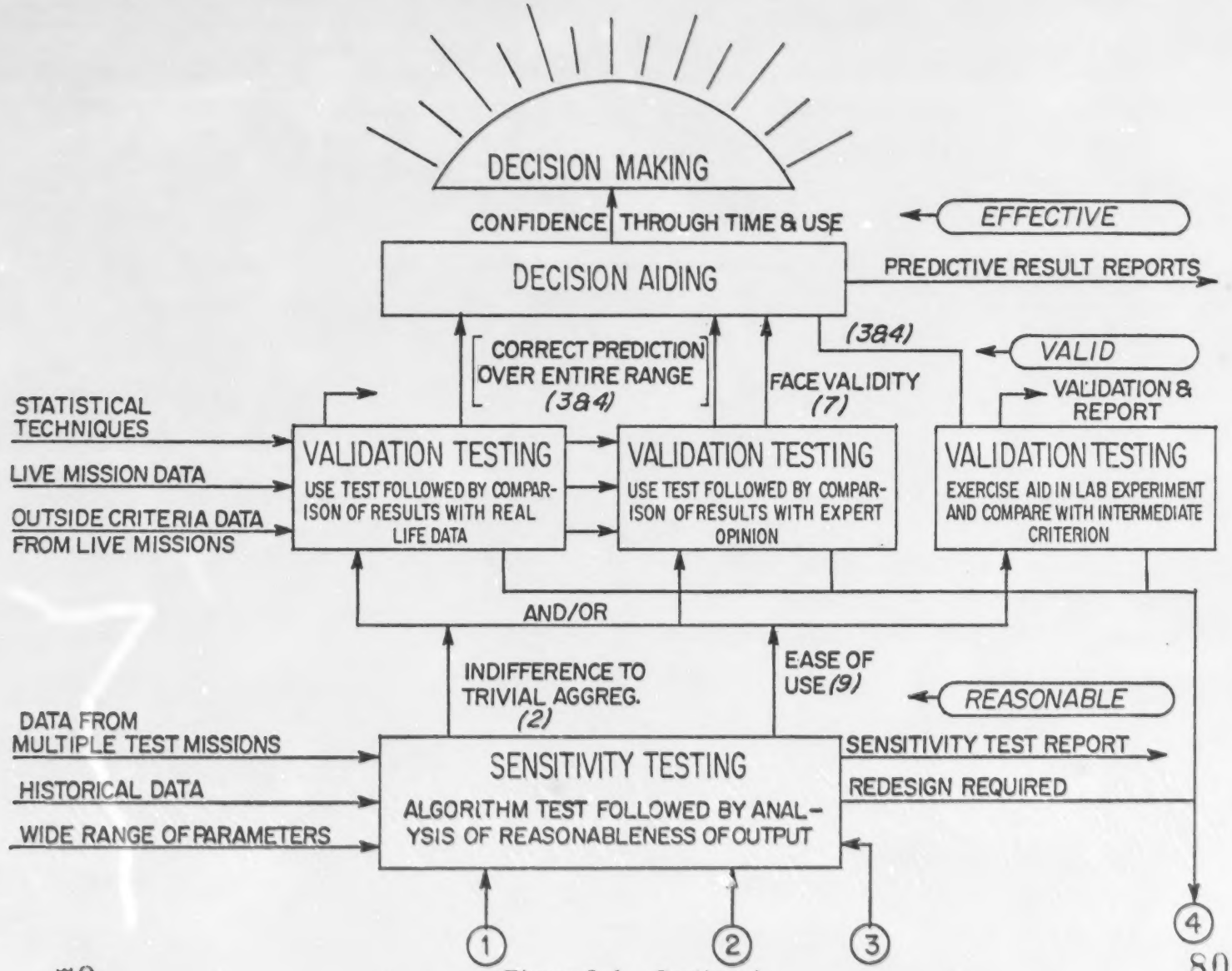


Figure 3-1. Continued

It is suggested that a model whose design meets the criteria emanating from the model design box be said to be "suitable" (see lowest oval). A model which is programmed and debugged enters a state here called "testable." After sensitivity testing, (and the implementation of corrections to the model as required), a model is said to be "reasonable." Following adequate validation testing, a model is termed "valid" or "useable" for decision aiding and, after the experience of use, a model is "operative," "proven," or "effective." The various types of data and information required as inputs to each phase are shown entering from the left with the resulting documentation outputs exiting to the right.

In the present study both the sensitivity testing and validation testing stages apply. Figure 3-1 shows that the three primary modes of validation testing are:

- compare model output against real mission criterion data
- compare model output against expert opinion
- compare model output against laboratory data

An alternate, expensive approach exists for these three modes of validation for the case in which real mission data acquisition is impossible or too dangerous. That is, the potential to develop two (or more) simulation models by independent groups of analysts. Validation can be accomplished then by comparing them against each other or against a single set of expert opinion relative to a preselected set of criteria. No example of such a case is currently known.

Model Validation

The model validation proceeded in several stages. First, a number of tests were designed to measure the operator parameters required for input to the model. These tests were administered to the personnel who employed the TWSEAS system during the field exercise on which the validation was based. Second, the TWSEAS performance during the field exercise was measured in order to collect the data required for the development of criterion information. Based on the input data describing the characteristics of the TWSEAS personnel who man the system during an actual field exercise and scenario for the exercise, the model was run to simulate the TWSEAS performance during the field exercise. Finally, the output data from the model and the TWSEAS criterion data from the field exercise were compared for similarity.

Individual Parameter Data Development

The initial step in the work involved participation in a briefing with the Marine Corps officers and noncommissioned officers who were scheduled to participate in the TWSEAS operation during an anticipated field exercise. In this briefing, the overall scenario and each participant's individual responsibilities were discussed. The group was also addressed by a representative who described the purposes and goals of the NETMAN validation. It was emphasized that the intention was to collect data to evaluate the NETMAN model and not to evaluate the performance of these individuals. The participants were informed that information concerning any given individual and his performance would remain completely confidential. It is believed that the acceptance and cooperation of all personnel was achieved. Then, the collection of individual data, for entry into the ultimate simulation of these persons, commenced. A total of 20 TWSEAS referees and radio operators was involved.

Task Performance Time for Referees

Within NETMAN, simulation of the performance of each task by the control system personnel simulated is partially based on the time it requires the personnel to complete the elements of the task. These data are called task analytic information within the model and are supplied as input information. Because simulation of the characteristics of the referees who participated in the field exercise was required, the basic task performance information required by NETMAN was collected from the same referees who participated in the actual exercise. During this data collection, the basic time required to process a message was measured. Each referee was given 21 scenarios, one at a time, and instructed to compose a TWSEAS message using the available information and the TWSEAS coding book. The scenarios consisted of the actual decoded format for the messages which constituted correct answers. Figure 3-2 shows an example of a problem scenario. The referee was asked to compose each message as accurately and as quickly as he could.

Figure 3-3 shows the data collection form used for recording the referee task performance information as required by NETMAN. The times were recorded on the data sheet in the form of cumulative elapsed time. Timing began as soon as the message was given to the referee. When the referee stopped leafing through the TWSEAS message book, the time was recorded under "Message type selection." If the referee changed his mind and moved to another message type, an "F" was recorded under "Success/Fail" and the time of final selection of a message type was recorded under "Touch up." If the

Scenario 1

You are assigned to unit 12 and decide to report the evacuation of casualties by motor transport. It is day D + 2 of the exercise and the time is 1045. Evacuation of the casualties was requested at 0945, and field medical cards were prepared for all casualties at that time. At 0800, one Marine was killed in action and 16 were wounded requiring evacuation. The casualty sorting was "above average" and evacuation procedures were effective.

Figure 3-2. Information provided to referees for composing one TWSEAS message.

APS Project NAME - Task Analytic Data Collection for Referee

Identification No. _____ _____ Date _____

Start time _____ _____ Message No. _____
(hours) (min.)

Stop time _____ _____ Observer _____
(hours) (min.)

	Duration (secs.)	Success/Fail	Touchup (secs.)
Message type selection	_____	_____	_____
Message encoding	_____	_____	_____
Visual check of entries	_____	_____	_____
Delivery to radio operator	_____	_____	_____

Comments:

Figure 3-3 Task analytic data collection form for referees.

referee was satisfied with his initial selection, an "S" was entered under "Success/Fail" and the "Touch up" time entry was left blank. If the referee selected a wrong message type, the error was not brought to his attention until the end of the complete data collection.

The time of writing the last digit of the TWSEAS code was recorded under "Message encoding." If the referee changed an entry, an "F" (otherwise "S") was entered under "Success/Fail" and the time of completing the final touchup correction was entered.

Similarly, the time of completion of the final check of the message was recorded under "Visual Check." This end point was taken to be the point at which the referee looked up from the message or the time at which the referee finished writing the time of message completion.

The timing ended when the referee returned the completed message and the scenario to the test administrator. This time was recorded under "Delivery to radio operator."

During the message preparation, the referee was not allowed to interupt for questions and the timing, once started, was carried through to completion. Any delay in the message handling process, such as dropping a pencil, was considered to be a possible occurrence in the field and was included in the timing.

Mean time, across referees, was calculated for each task element of each message type. The message types are described in Table 3-13. The resulting means and standard deviations are shown in Table 3-5. Currently, NETMAN is limited to a total of eight task analyses in one simulation. Since at least one task analysis must be assigned to each level of the system, three are accounted for by the controller, computer, and radio operator. This allowed an allocation a maximum of five task analyses for use by referees. Message types 5 and 2, as shown in Table 3-5, were omitted from this simulation. This omission seems justified on the basis that no messages of type 5 and only two messages of type 2 were ultimately involved in the field exercise. Utilization of "Success/Fail" data is presented below in the discussion of precision determination (see Table 3-9). Task elements 1 through 4 in Table 3-5 correspond to the four tasks in Figure 3-3.

Table 3-5

Referee Task Analyses

Message Type	Task Element	Mean Task Duration (secs.)	Standard Deviation (secs.)
1	1	45.0	40.0
	2	117.1	41.4
	3	5.8	9.5
	4	3.1	3.2
2*	1	37.8	23.0
	2	137.5	74.9
	3	5.2	9.6
	4	3.0	3.9
3	1	39.4	21.8
	2	135.4	70.1
	3	5.7	9.4
	4	2.2	2.8
4	1	44.9	18.8
	2	203.2	75.9
	3	2.4	4.2
	4	3.1	3.4
5*	1	43.0	29.9
	2	161.7	43.2
	3	2.6	3.4
	4	2.3	2.6
6	1	38.2	14.3
	2	153.3	58.2
	3	6.0	12.5
	4	7.0	10.7
7	1	34.9	10.5
	2	182.3	39.5
	3	5.6	12.6
	4	4.7	7.4

* Omitted from simulation.

Task Performance Time for Radio Operators

NETMAN requires the same type of task analytic data for radio operators as for referees. The procedure for collecting task element performance time relative to radio operators paralleled the procedure followed for referees. Each radio operator was sequentially given 21 messages and was asked to transmit the message using a DMED (a message transmittal device) box. As for the referees, the time to complete various message segments, as required by the NETMAN, was recorded for the radio operator using the form shown in Figure 3-4.

Time was recorded in the form of cumulative elapsed time. Timing started when the radio operator first received the message. When the radio operator put the message down or placed it in position on the top of the DMED box, as some of them did, this time was recorded under "Visual inspection of message." If the radio operator decided to revise the message, an "F" was entered under "Success/Fail" and the revision time was entered under "Touch up." When the radio operator finished dialing the final digit of the 22 digit TWSEAS message, the time was recorded under "Visual check." If the radio operator went back to change any dialed in digit, a failure was recorded and the final time for completion was recorded under "Touch up."

After setting up the DMED box, the radio operator set the radio to the proper transmittal frequency. The time of completion of radio set up, usually indicated by reaching for the headset, was recorded under "Radio set up." After setting up the radio, the radio operator pressed a button to transmit the message and then waited for a beep signal which indicated acknowledgment by the computer. In actual operation, a message is transmitted and acknowledged twice. However, there was no actual computer acknowledgment during this data acquisition period. The radio operators were told to wait a few seconds--as if they were waiting for an actual acknowledgment. The time at which they completed the transmittal was recorded under "transmittal."

The final task to be performed by the radio operator was recording the message information in his log book. The time of completion of this recording was entered under "Storage of message" and this entry represented the final time of the message processing.

In the case of the radio operator, only one task analysis was used in the ultimate simulation. The mean and standard deviation

AFS Project NAME - Task Analytic Data Collection for Radio Operator

Identification No. _____ Date _____

Start time _____ (hours) _____ (min.) Message No. _____

Stop time _____ (hours) _____ (min.) Observer _____

	Duration (secs.)	Success/Fail	Touchup (secs.)
Visual inspection of message	_____	_____	_____
Visual check	_____	_____	_____
Radio set up	_____	_____	_____
Transmittal	_____	_____	_____
Storage of message	_____	_____	_____

Comments:

Figure 3-4 Task analytic data collection form for radio operators.

of each task element, as derived from the above described test data, is shown in Table 3-6. These times and standard deviations were used during the validation simulations for all message types.

Other input information required by NETMAN for each person included in the simulation included the stress threshold, level of aspiration, a speed (proficiency) factor, and a precision factor. The basis for these data is described below.

Stress Threshold Determination

Within NETMAN, an individual stress threshold is employed to affect the success experience and response time of each person simulated. To derive the stress threshold for each individual, a mirror tracing task was employed. This task, due to the fact that the visual feedback is disoriented by the mirror reversal, is stress producing for most people. Stress was measured through use of electrodes attached to the fingers of the nonpreferred hand while the preferred hand was used in the mirror tracing task. This test required approximately 20 minutes to administer.

Skin conductance was measured through the use of an Autogenics Model 3000 Skin Conductance Dermograph. After a practice trial, the conductance level change was measured throughout a formal test.

The maximum change in skin conductance level was converted to a stress threshold for each referee and radio operator through use of the following equation:

$$STRM = -.0909\ STR + 2.9091$$

Where STRM = stress threshold

STR = stress level = change in skin conductance level

This equation was derived by translating the range and direction of conductance level changes derived from the test data to the stress threshold range as scaled within the model. The result was a linear transform with a high threshold reflecting a low conductance change. The resulting stress thresholds are shown in Table 3-7.

Table 3-6

Radio Operator Task Analytic Time

Task Element	Mean Task Duration (secs.)	Standard Deviation (secs.)
Visual Inspection	8.3	7.4
Visual Check	43.0	14.0
Radio Setup	7.6	7.3
Transmittal	11.1	6.1
Storage	38.0	11.3

Table 3-7

Stress Threshold of Each Referee and Radio Operator

Subject Identification Number	Skin Conductance Level Change	Stress Threshold
	Referees	
1	8.10	2.17
2	3.83	2.56
3	4.58	2.49
4	9.67	2.03
5	7.52	2.23
14	4.79	2.47
15	5.12	2.44
16	8.83	2.11
17	10.00	2.00
18	4.75	2.48
	Radio Operators	
6	6.92	2.28
7	7.75	2.20
8	3.00	2.64
9	8.58	2.13
10	3.18	2.62
11	4.33	2.52
12	5.67	2.39
13	3.83	2.56
19	2.96	2.64
20	10.00	2.00

Level of Aspiration Determination

Level of aspiration is a variable within the NETMAN which affects each simulated person's response to success or failure. Level of aspiration is generally an indication of an individual's need for achievement. It determines the level of success which he anticipates and strives towards. Within the model, the effect of level of aspiration is a somewhat complex function of the stress level present, the individual's "Success/Fail" history, and his initial level of aspiration. The result is an effect on success probability.

To obtain an estimate of the level of aspiration of each of the persons to be simulated, each referee or radio operator was told that most persons can mark 30 "x's" in a matrix of square boxes in 30 seconds. He was then asked how many he could do in 30 seconds.

To derive the required level of aspiration, the following equation was employed.

$$\text{Level of Aspiration} = \frac{\text{Estimate}}{30}$$

The upper limit of level of aspiration is set at 1.0. Table 3-8 presents the resulting level of aspiration indices.

Individual Precision Determination

Operator precision is another individual input required by the NETMAN model. Precision is used within the model to affect the number of mistakes which are made by a simulated operator. Table 3-9 shows the number of errors made by each referee and radio operator in the message preparation (described earlier). The smaller number of errors made by radio operators is probably a function of the more basic task performed by radio operators. The mean number of errors was calculated as a percentage of the number of possible errors. In the case of the referees, there were four task elements which might be failed and 21 messages. This means that there were 84 possible errors and, since 11 errors resulted on the average, the overall probability of success across all referees was .869. For the radio operators, there were five task elements resulting in 105 possible errors. The average radio operator success probability was .984. The success probability of each man was calculated in the same way. The individual precision factors were calculated as shown below.

Table 3-8

Level of Aspiration of Each Referee and Radio Operator

Subject Identification Number	Referee	Level of Aspiration
1		1.00
2		1.00
3		1.00
4		1.00
5		.67
14		1.00
15		1.00
16		1.00
17		1.00
18		1.00
	Radio Operator	
6		1.0
7		1.0
8		1.
9		1.
10		.
11		1
12		
13		
19		
20		

For individuals with fewer mistakes than the average:

$$PREC = 1 - \left[(PROB - AVPRB)/[5\ (1 - AVPRB)]\right]$$

For individuals with more mistakes than the average:

$$PREC = [6 - (PROB/AVPRB)]/5$$

Where PREC = individual precision factor
PROB = individual success probability
AVPRB = mean group success probability

These equations are the same as the equations which determine the effects of precision within the model. Precision within the model is scaled the same as speed. A 1.0 precision level is the nominal condition and has no effect on success probability. A precision level of .9 produces a higher success probability and fewer errors. A precision level of 1.1 produces reduced success probabilities and more errors.

The resulting precision factors are shown in Table 3-9.

Individual Speed Factor

Within NETMAN, performance time for various task elements is partially based on a speed factor for persons simulated. This speed factor is another required input.

The speed factor input for each referee and radio operator was calculated on the basis of individual message processing time in comparison with the overall message processing time for the respective group. Table 3-10 shows the mean time for each referee and each radio operator to process a message. These times are means across the 21 messages processed by each man. The time for the referees ranged from a low of 118.2 seconds to a high of 261.4 seconds. The mean time to process a message across all referees was 187.5 seconds. The formula used to compute each referee's speed factor, F, was:

$$F = \frac{\text{Individual Mean Message Processing Time}}{\text{Group Mean Message Processing Time}}$$

For the radio operator, processing time ranged from 85.1 seconds to 130.8 seconds with a mean of 108.7 seconds. The speed factor for each radio operator was calculated in the same manner as for referees.

Table 3-9

Success Probability and Precision of Each Referee and Radio Operator

Subject Identification Number	No. of Errors	Success Probability	Operator Precision
	Referees		
1	1	.988	.818
2	9	.893	.963
3	9	.893	.963
4	10	.881	.982
5	12	.857	1.003
14	24	.714	1.036
15	7	.917	.927
16	17	.798	1.016
17	14	.833	1.008
18	7	.917	.927
Average	11.0	.869	1.000
	Radio Operators		
6	1	.990	.925
7	5	.952	1.007
8	0	1.000	.800
9	0	1.000	.800
10	0	1.000	.800
11	4	.962	1.004
12	1	.990	.925
13	3	.971	1.003
19	2	.981	1.001
20	1	.990	.925
Average	1.7	.984	1.000

Table 3-10

Mean Messages Processing Time and Speed Factor for Each Referee and Radio Operator

Subject Identification Number	Mean Processing Time (secs.)	Speed Factor
	Referees	
1	185.3	.99
2	207.0	1.10
3	132.4	.71
4	186.1	.99
5	205.3	1.09
14	220.8	1.18
15	134.9	.72
16	261.4	1.39
17	223.2	1.19
18	118.2	.63
Average	187.5	
	Radio Operators	
6	110.3	1.01
7	126.3	1.16
8	98.2	.90
9	115.3	1.06
10	105.5	.97
11	111.0	1.02
12	130.8	1.20
13	85.1	.78
19	106.2	.98
20	97.9	.90
Average	108.7	

The resultant speed factor, for each referee and radio operator is shown in Table 3-10.

Background Information

All personnel tested were asked to complete a background data form. The purpose of this form was to collect information about the experience and background of the referees and the radio operators involved in the field exercise. The referees included five first lieutenants and five sergeants. Their time in the Marine Corps ranged from 31 months to 174 months. The mean service time in the Marine Corps was 84 months. The level of education was four years of college for all of the lieutenants and high school graduate or equivalent for the sergeants. Only one of the referees, a sergeant, had worked with the TWSEAS system before.

The radio operator group consisted of one sergeant, seven corporals, and two privates, first class. Their Marine Corps service ranged from 18 months to 102 months with a mean of 38 months. Eight of the 10 were high school graduates; one did not graduate from high school, and one had one year of college. Only one radio operator, a private first class, had worked with the TWSEAS system before.

Message Frequency and Frequency of Message Type

Message frequency is a nonpersonnel input information requirement of the model. This input is the mean number of messages generated per referee per hour and its standard deviation. It is presently required that these numbers be entered into the model in integer form (i.e., no decimal point). Table 3-11 shows the number of TWSEAS messages which were actually tracked during the field exercise of the validation process.

These data were employed to provide the input information necessary for simulating the field exercise message frequency. The frequency of each message type in the field was computed so that this required input information could be used in the model's simulation of the field exercise. Table 3-12 shows the actual frequency of each of the listed TWSEAS message types over two days of the field exercise. These data formed the basis for the message type input requirement of the simulation aspect of the validation process.

Table 3-11

Field Message Frequency per Hour

Hour	Day 1	Day 2
1	4	4
2	2	3
3	4	7
4	0	4
5	1	4
6	2	4
7	8	4
8	3	3
9	3	3
10	1	1
Total	28	38

Table 3-12

Message Type Frequency

TWSEAS Message Type	Frequency	Description
7	1	Scenario Event Report
10	1	Personnel Replacement Report
11	11	Personnel/Equipment/Small Arms Ammo Initial/Update
30	31	Situation/Location Report
31	4	Planned Move Report
32	5	Contact Report
35	3	Attach Plan Report
36	4	Defense Status Report
38	2	Attachment/Detachment Report
41	1	Casualty Assessment Report
42	1	Casualty Evacuation Report
80	2	Assault Wave Landing Report

To reduce the number of message types to the model's capacity, the number of message types was collapsed as shown in Table 3-13, which shows the correspondence between the 12 TWSEAS message types (Table 3-12) and the five message types used in this model application.

Computer Runs Performed

The input data, as derived in the prior sections, were entered in the required format for simulation within the NETMAN model. One run was organized to simulate TWSEAS operation during the first day (D - day) of a battalion level field exercise and another run was completed to simulate D - day plus one. Ten hours of operation were simulated each day to provide a total of 20 hours of model predictions which could be compared with TWSEAS performance.

Observer Training

Seven Marine Corps observers collected the required TWSEAS performance information. The group was composed of four Marine Corps' Second Lieutenants, one First Lieutenant, and two retired Marine Corps' Sergeants Major.

The observers were trained during the two days prior to the exercise. The training included a briefing on the overall objectives of the present work, practice in filling out TWSEAS messages, and practice in timing simulated referee tasks and radio operator tasks. The observers completed the same 21 TWSEAS messages which were used to test the referees. The results were reviewed and errors were corrected. The observers were instructed on the behaviors which signaled various aspects of the message processing. For example, referee message processing started when the referee exhibited an initial behavior which involved the message. Frequently, picking up the TWSEAS message code book was the starting point; but, sometimes, a referee filled out part of the message, such as unit identification and time, before picking up the TWSEAS booklet. The observers were also instructed to comment on any unusual occurrences.

After thorough discussion of the behaviors which formed the basis for timing and how to complete the data collection form, the observers received practice in form completion in a simulation situation. During the simulation, the observers recorded the appropriate data. Approximately 10 messages were involved. The data forms, as completed by each observer, were individually reviewed for accuracy and completeness. Further instructions were given, as required, until each observer was proficient in all aspects of the data collection.

Table 3-13

Consolidation of Message Types

Message Type	Description	Consolidated To Type
30	Situation/Location Report	1
41, 42	Stream messages - request and delivery.	2
7, 31, 32	Stream messages - ground engagement.	2
35, 36	Other messages requiring controller action and/or complex encoding by the referee, with a high probability of occurrence.	3
None	Other messages requiring controller action and/or complex encoding by the referee, with a low probability of occurrence.	Not Used
10, 11	Other messages requiring no controller action and normal encoding by the referee, with a high probability of occurrence.	4
80, 38	Other messages requiring no controller action and normal encoding by the referee, with a low probability of occurrence.	5

Field Exercise

The logic for the validation process called for comparing the results of a field exercise with the NETMAN predictions of these results when the performance of the same personnel that are included in the field exercise is simulated. Prior sections described the collection of the required personnel descriptive data for the model. The present section describes the field exercise which provided the basis for the results comparison.

In the field exercise, a full Marine Corps battalion landing team assaulted a beach with armor and mechanized units. The objective of the exercise was to rescue an abducted member of congress. The details of the exercise included numerous searches, engagements, and movements. The exercise started with an assault on a beach and terminated with the withdrawal of troops back to the ships. The exercise was observed for two successive days (10 hours each day). The withdrawal took place on the third day and was not observed or simulated.

Field Exercise Data Collection

Two types of field exercise data were collected. The first type was provided by observers who traveled with mobile units and collected data concerning referee and radio operator message processing. The second type of field data was collected at the control van and concerned the computer and controller message processing.

The observers assigned to the control van were one psychologist and one ex-army Captain.

Figure 3-5 shows the field data collection form used by all observers to record data. The identification number was used to relate test data to field data. Since the same form was used for referees, radio operators, and controllers, the position entry was required. The TWSEAS message type and the time entered on the message were entered to allow tracing the message through the system -- from the referee through the control van. The times entered were the message start and stop times. The delay for transmission was entered only for radio operators. Comments of all types were included to promote greater insight into the functioning of the system.

To provide a measure of overall effectiveness, the controllers were asked to rate the overall system information handling effectiveness

APS Project NAME - Field Data Collection Form

Identification No. _____ Date _____

Position (R, RO, or C) _____ Message No. _____

Unit _____ Observer _____

Message Type (TWSEAS) _____

Message Time Tag _____

Clock Time

Start _____ _____ Stop _____ _____
(hours) (min.) (hours) (min.)

Elapsed Time _____ Delay for Transmission _____
(sec.) (sec.)

Comments:

Figure 3-5 Field data collection form.

on a magnitude estimation scale. Reference points and an example were provided on the effectiveness rating sheets.

Discussion of Field Exercise Data

When selecting the exercise for use as a validation exercise, the representativeness and appropriateness of the exercise were carefully considered. While the field exercise was fully realistic, the status reporting system (TWSEAS) was limited by radio failures, insufficiently trained (in the use of TWSEAS equipment) personnel, variations in message handling procedures, and the relatively small number of messages measured in their entirety.

Radio problems, while sending the TWSEAS messages, provided many delays, changes in procedure, or attempts to solve these problems. Radio operators changed batteries, changed antennas, moved their radios to other locations, and used the voice frequency to verbally relay the TWSEAS messages. Often, the temporary radio transmission problems caused the almost complete cessation of attempts to send TWSEAS messages. Other techniques, such as messengers and voice relays, were substituted. These non TWSEAS techniques are not simulated or accounted for in the present structure of the NETMAN model.

Additionally, deviations from prescribed procedures occurred on a number of occasions. For example, occasionally a referee would compose a message on the DMED box and transmit it himself. Such deviations in procedure are not simulated within the NETMAN.

Such problems are to be expected in any quantitative data collection effort which takes place in a real field situation. Nevertheless, the affect of such problems on criterion data quality may not be trivial.

Validation Analysis

The agreement between the model's predictions and the TWSEAS performance during the field exercise was examined according to the eight criterion variables described earlier. Agreement was measured for the following measures as computed by the model: thoroughness, responsiveness, overall effectiveness, total message processing time, referee message processing time, radio operator message processing time, controller message processing time, and transmission delay. These eight aspects will be examined individually in turn.

Thoroughness

Thoroughness within the model is calculated as the number of messages completely processed divided by the number of messages available to be processed in each hour. The criterion data were highly variable and responsive to factors which are not simulated within the model. The data from the field were therefore considered to be a range of data into which the model's output should fall. A tally was made of the number of times that the model predictions fell within one standard deviation of the field data or within one half standard deviation of the field data. There were 20 comparisons based on 10 hours per day for two days. Out of the twenty comparisons, nineteen of the models hourly predictions fell within .5 standard deviations of the field hourly data. According to the sign test (two tailed), this agreement ratio (19 out of 20) is statistically significant at the .001 level.

Since the thoroughness index is fundamentally a measure of network throughput, this finding seems particularly important. System planners and evaluators are interested in throughput because of the obvious relationship of throughput to system efficiency.

Responsiveness

Responsiveness, within the NETMAN model, is calculated as the ratio of human message processing time to total message processing time. Human message processing time includes the sum of the referee, radio operator, and controller time spent working on a message. Total message processing time includes human time plus other time such as delays or waiting time. Model responsiveness predictions fell within one standard deviation of the actual TWSEAS hourly performance data 18 times out of 20. This agreement ratio is indicated to be statically significant by a two tailed sign test at the .001 level of confidence.

The indication of validity for this responsiveness index will be of value to system designers, planners, and users because it provides a basis for knowing whether or not the human is contributing most to system throughput time. In a sense, the responsiveness index complements the thoroughness index. The latter provides a time oriented measure while the former provides a message load oriented index. Quite obviously, to gain insight into the reasons for a high or a low responsiveness index, the user will want more detailed, valid information on individual processing time. The validity of such data, as yielded by the model, is discussed later under the Referee Message Processing Time, Radio Operator Processing Time, and Controller Message Processing Time headings.

Overall Effectiveness

Overall effectiveness, within the model, is calculated as a composite of four individual effectiveness components -- thoroughness, accuracy, completeness, and responsiveness. The relative weights for each component are entered by the model's user. In the simulation of the TWSEAS, equal weights were assigned to each component. The overall effectiveness equation also takes into account the correlations between the effectiveness components. The values entered to describe these intercorrelations were all 0.5.

For the TWSEAS performance criterion, the opinion of the controllers was used as the measure of overall effectiveness. One problem with this criterion is that the controllers based their evaluations on the performance of the total complex while the NETMAN simulation only included those messages which were successfully transmitted using the normal TWSEAS system.

The NETMAN model provides hourly data for overall effectiveness. The model's predictions were compared with plus or minus one standard deviation of the TWSEAS criterion in a two tailed sign test. The result was not statistically significant. The controller evaluations of overall system effectiveness were substantially lower than the model's prediction. This lower rating by the controllers may be due to the presence of equipment and radio transmission difficulties, which were bypassed through the use of alternative delivery methods. These difficulties were not simulated in the model. Only messages which are able to pass through the normal TWSEAS channel are considered in the simulation.

Alternatively, the weights which were assigned to the four effectiveness components may have not adequately reflected the subjective values of the controllers. Or, the assigned intercorrelations may have been erroneous.

Total Message Processing Time

Total message processing time begins when a referee starts working on a message and ends when the controller finishes working on the message. Total message time includes referee working time, radio operator working time, and controller working time as well as all delays and interruptions which occur between working times. The measurement of time required to process a message completely is possibly the best single measure of system performance. Total message processing time was obtained for each hour of the TWSEAS operation. Total message processing time is also an output function from the NETMAN model. The TWSEAS criterion data were compared with the NETMAN output by means of a matched pairs t-test. The resulting t value was .381. With an N of 20, this value is not statisically significant (i.e., the model versus TWSEAS time difference was not significantly different). This finding supports a contention of reasonable correspondance between predicted and actual TWSEAS total message processing time. As a result, confidence in the model's ability to predict total message processing time is significantly increased; and, since total message processing time is an especially important aspect of the simulation, confidence in the use of the NETMAN model can be increased.

Operator Processing Time

During the field exercise, the time spent working on each message was measured for each individual (of those who were under observation). These data were averaged across operator type to produce mean time per message per hour for each operator type (referee, radio operator, or controller). The NETMAN produces parallel data. Independent comparisons were performed for referee processing time, radio operator processing time, and controller processing time. In each comparison, the frequency of the NETMAN prediction data falling within plus or minus one standard deviation of the mean of the TWSEAS observational data was determined. In each case, 20 out of 20 of the model's predictions fell within one standard deviation of the TWSEAS criterion. For each position, the two tailed sign test was statistically significant at the .001 level of confidence. This finding indicates considerable agreement between the NETMAN's predictions and the TWSEAS time criterion at all levels of message processing. The agreement between predicted and criterion time is especially important since modeling manipulations of operator time (for example, by varying operator proficiency level or through task revision) can provide important information to system designers concerning the required number of personnel and the required training of personnel in new systems. Such simulations

would also provide information concerning the workload imposed on personnel, and, therefore, the availability of such personnel for additional responsibilities. Finally, the individual time data may now be employed as a basis for gaining additional insight into the responsiveness measure, as described above.

Transmission Delay

Transmission delay in the TWSEAS is represented by the time between a first attempt to transmit a message and the time recorded by the computer as the time of message entry. In a "perfect" communication network, this time would be nil. Within NETMAN, transmission delay is defined similarly. Transmission delay can occur for a variety of reasons. Equipment problems are one of the most common. Another reason for delay is that a line of sight radio transmission frequently induces difficulty when used under less than ideal conditions. A third, less prevalent, reason for transmission delay in TWSEAS is line or channel congestion.

Only one overall transmission time is available per day from the NETMAN computer model. This one prediction time was compared with the data for each hour from the TWSEAS criterion. Comparison of the predictions of the NETMAN model with the TWSEAS performance indicated that the model data were within .5 standard deviations of the criterion data in 19 out of the 20 comparisons. This agreement is statistically significant (sign test) at the .001 level. Such a finding is consistent with the agreement between the model's prediction and overall processing time in the TWSEAS. Since transmission delay is an ingredient of overall message processing time, one would not expect agreement in the overall unless the subaspects also agree.

The NETMAN model does not simulate equipment problems as such. NETMAN only simulates the resulting time delays. Accordingly, the obtained level of agreement is especially encouraging.

Integration of Validation Data

Figure 3-6 shows the correspondence between the NETMAN model's predictions and the TWSEAS data in terms of deviation from the TWSEAS data. The "x's" represent the mean model predictions across all hours, while the mean and standard deviation of the horizontal axis represent the mean and standard deviation of all field data points. Two of the predictions agreed almost exactly with the respective criterion--thoroughness and total processing time. Five of the predictions --responsiveness, referee processing time, radio operator processing time, controller processing time, and transmission delay,

were within one standard deviation of the mean of the field data. Only one of the indices investigated, overall effectiveness, was more than one standard deviation from the mean of the field data. In this case, the deviation exceeded two standard deviations. As was indicated, there was a lack of correspondence between the TWSEAS overall effectiveness as rated by the controllers and overall effectiveness as measured by the model.

Table 3-14 summarizes the statistics and results of the eight validation analyses. The sign tests were all structured to test the degree of agreement between model and field data. Agreement to within one standard deviation was found in all cases except for overall effectiveness. A t-test was performed to compare model and field total processing time. No statistically significant difference was found between model predictions and TWSEAS data in this case.

Discussion

The results of the validation support contentions favoring the validity of the NETMAN model relative to all but one of the diverse TWSEAS criteria employed. The degree and magnitude of the diverse events involved in field exercises in which a TWSEAS like system is involved brings out clearly the need for computer models in this area. Where variable occurrences overshadow design variables, the need for improvements is very hard to evaluate. Each field exercise, by its nature, is expensive and provides only a small sample of possible conditions. Within a model, however, the effects of such events can be assessed by combining the effects of many low probability events through stochastic iteration.

The criterion quality is believed to be reasonably acceptable. One problem was the limited amount of data, i.e., the number of messages that the referees were able to transmit from the field over the normal TWSEAS system. The small message sample size represents the primary limitation on the adequacy of the criterion data.

The measurement of the TWSEAS performance, as performed by the field observers, is believed to be highly reliable. Although no formal independent measures were available, the use of discrete time points and message identifiers when coupled with the training that was administered, allow a high degree of confidence in the reliability of the data. The exercise data, as collected, were highly objective. Discrete, predefined starting and stopping points were identified for time measurements. Subjective evaluations were used only in the case of the controller's overall effectiveness evaluation.

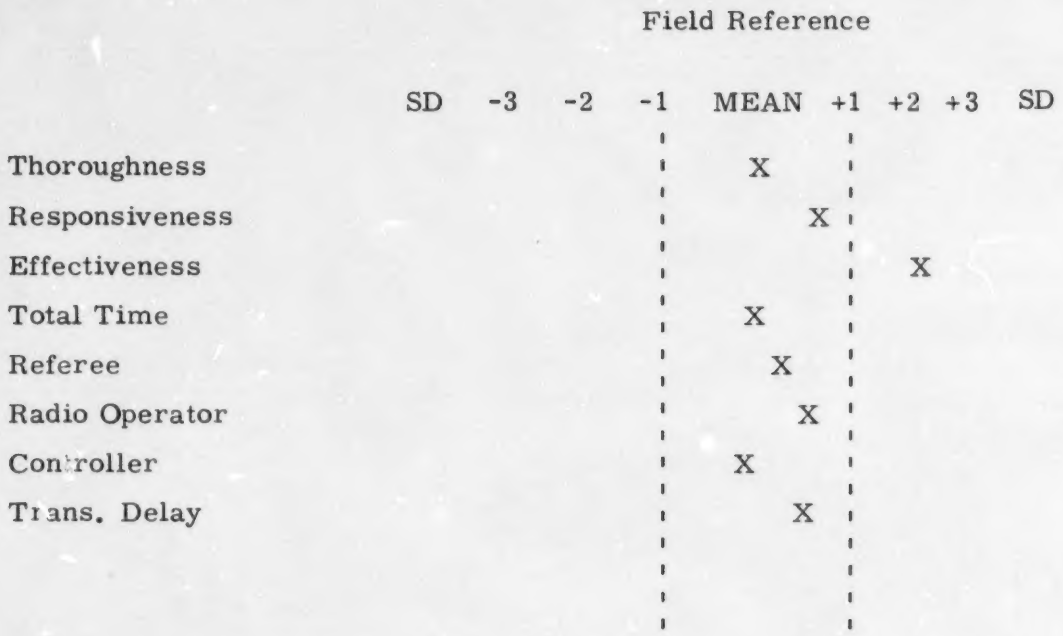


Figure 3-6. Model predictions versus field data.

Table 3-14

Summary of Validation Statistics

Variable	Test	Not Significant	Agreement Achieved .01
Thoroughness	Sign		✓ **
Responsiveness	Sign		✓ *
Overall effectiveness	Sign	✓	
Total message prccessing time	T-test	(✓)***	
Referee message processing time	Sign		✓ *
Radio operator message processing time	Sign		✓ *
Controller message processing time	Sign		✓ *
Transmission delay	Sign		✓ **

* Within 1 sigma

** Within .5 sigma

*** The t-test as used is a measure of difference while the Sign test as used is a measure of agreement. A non statistically significant difference indicates agreement between the model's prediction and the criterion data

TWSEAS performance data required little or no transformation, rescaling, preprocessing, or translation in order to be used as criteria. The calculations and processing described ir the early part of this chapter were involved in the calculation of input data for the model, not for the transformation of criteria data.

The high variability of the TWSEAS performance, confounded as it was by radio and other equipment problems greatly limits its sensitivity to operator performance factors, fatigue and other factors to which the model is sensitive. The higher sensitivity of the model to these effects is one of the primary advantages of the model.

IV. SUMMARY AND CONCLUSIONS

Extensive testing of the NETMAN computer model was completed. This testing included about 60 different simulations for sensitivity test purposes and one simulation of a real system for validational purposes. In the sensitivity tests, the effects of a variety of personnel variables, workload variables, manpower configurations, and task variables were examined for consistency, reliability, independence from trivial effects, and rationality. In general, the results were found to be reasonable, appropriate, and useful. The most important variables were operator speed, operator precision, and network configuration. The psychological factors, stress threshold and level of aspiration, exerted a much less powerful affect on output.

Another result of the sensitivity tests was an evaluation of the ease of use and the cost of use of the NETMAN computer model. In terms of cost of use, the NETMAN program was found to be very efficient. In most cases, an extensive number of iterations is not required. The findings relative to ease of use were equally satisfactory. Setup for a new simulation is now estimated as not more than 40 man hours, for an experienced analyst. This time involvement will depend, to some extent, on the complexity of the system being simulated. Once a simulation has been organized, however, changes and rerunning the simulation are readily accomplished.

The test vehicle chosen for evaluating the NETMAN model's validity was the Marine Corps' Tactical Warfare Simulation, Evaluation and Analysis System (TWSEAS). Message processing information on eight different aspects of the system performance was collected to serve as the criteria in the validation analysis. Personnel working in the TWSEAS were pretested to provide the required personnel input data for the NETMAN model. The predictions from the simulation were compared with the criterion data during actual TWSEAS use. The model's predictions were within plus or minus one standard deviation of the criterion data 96 percent of the time for "thoroughness," 90 percent of the time for "responsiveness," and 100 percent of the time for referee message processing time, radio operator message processing time, and controller message processing time. On the negative side, the model's prediction of overall system effectiveness was over two standard deviations from the mean of the "overall effectiveness" criterion measure. The discrepancy in this case is believed to be due to the controllers who provided the TWSEAS "overall effectiveness" criterion measure. The controllers evaluated the entire system while

in the model overall effectiveness is based on messages that are processed in the normal TWSEAS condition. Radio problems during the validation data collection often caused messages to bypass the normal radio link and take other routes to the control van. The radio difficulties and indirect message routes caused the controller to rate overall system effectiveness lower than was predicted by the model.

Overall, the results of the sensitivity and validation tests suggest substantial confidence in the NETMAN simulation model. However, a number of areas were identified for model improvement. Presently, the model computes means over all iterations but does not compute the standard deviation or variance of the data. Variance data would be useful in evaluating model output. Among other things, it would provide a way to estimate the number of iterations which are necessary to obtain model output stability.

At the outset of the current effort, a panel of experts was convened to discuss the present status of the model. A number of areas for model improvement were identified. Increasing the ease of use of the model was considered to be especially important by the panel. A number of factors which must presently be computed manually by the analyst preparing the input data for the model could be computed automatically as part of the computer program. As an example, the input data for message type frequency must be entered in the form of cumulative proportions. The model could be changed so that actual frequencies are entered with the model itself calculating the cumulative proportions. Similarly, there are a number of types of error which might be made in the assembly of input data. The program could be extended to check for many of these errors before simulation execution. As an example, when two networks are specified in the input, the number of men in each (some number greater than zero), must be specified. The program could very easily be modified to check that the entry in both cases is nonzero.

The NETMAN computer program already allows a number of changes to be made to the input data in an interactive mode. After calling the program from a terminal, the user is able to select from different categories of input, make changes, and then initiate a new simulation. This interactive mode is already useful but a number of additional methods for prompting and assisting the model user in this mode might be made available. There are also a number of variables in the model which cannot currently be changed in the interactive mode. Capability might be expanded in this regard.

The following conclusions appear warranted:

1. The NETMAN model is adequately sensitive over a wide range of parametric variation.

2. The parametric variation introduced during the current sensitivity tests produced results which are reasonable and which possess proper directionality.

3. The results of the validity tests indicate that NETMAN can provide information about exercise control system operation comparable to that available in a field exercise environment such as TWSEAS.

4. Several areas for model improvement are indicated.

REFERENCES

Campbell, D.T. & Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, *56*, 81-105.

Emshoff, J.R. & Sisson, R.L., *Design and Use of Computer Simulation Models*, London, MacMillan Co., 1970.

Hermann, C. Validation Problems in Games and Simulations, *Behavioral Science*, 1967, *13*, pp. 216-30.

Leahy, W.R., Lautman, M.R., Bearde, J.L., & Siegel, A.I. *A digital simulation model of message handling to the tactical operations system. III. Further extensions of the model for increased interaction.* Research Memorandum 74-11. U.S. Army Research Institute for the Behavioral and Social Sciences, August 1974.

Leahy, W.R., Siegel, A.I., & Wolf, J.J. *A digital simulation model of message handling in the tactical operations system. IV. Model integration with CASE and SAMTOS.* Wayne, Pa.: Applied Psychological Services, November 1975.

Leahy, W.R., Siegel, A.I., & Wolf, J.J. *A digital simulation model of message handling in the tactical operations system. V. User's guide to the integratal MANMOD/CASE/SAMTOS computer simulation.* Wayne, Pa: Applied Psychological Services, November 1975.

Siegel, A.I., Lautman, M.R., & Wolf, J.J., *A Multimethod-Multitrait Validation of a Digital Simulation Model.* Wayne, Pa.: Applied Psychological Services, 1972.

Siegel, A.I., Leahy, W.R., & Wolf, J.J. *A computer model for simulation of message processing in military exercise control and evaluation systems.* ARI Technical Report TR-77-A22. U.S. Army Research Institute for the Behavioral and Social Sciences, October 1977.

Siegel, A.I., & Wolf, J.J. *Techniques for Evaluation Operator Loading in Man-Machine System: Application of a Previously Derived Model to the Launching of an Air-to-Air Missile.* Wayne, Pa.: Applied Psychological Services, 1959.

Siegel, A.I., Wolf, J.J., & Leahy, W.R. *A digital simulation model of message handling in the tactical operations system. I. The model, its sensitivity, and user's manual.* Research Memorandum 73-5. U.S. Army Research Institute for the Behavioral and Social Sciences, November 1972.

Siegel, A.I., Wolf, J.J., Leahy, W.R., Bearde, J.L., & Baker, J.D. *A digital simulation model of message handling in the tactical operations system. II. Extensions of the model for interacting with subject and experimenters.* Research Memorandum 73-6. U.S. Army Research Institute for the Behavioral and Social Sciences, December 1973.

Siegel, A.I., Wolf, J.J., & Sorenson, R.T. *Techniques for Evaluating Operator Loading in Man-Machine System: Evaluation of a One or Two-Operator System Evaluative Model through a Controlled Laboratory Test*. Wayne, Pa.: Applied Psychological Services, 1962.